LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-615272

Understanding Cytochrome P450 Regioselectivity for Acetaminophen Metabolism

Y. Yang, S. E. Wong, F. C. Lightstone

February 1, 2013

Journal of American Chemical Society

Disclaimer

Understanding Cytochrome P450 Regioselectivity for Acetaminophen Metabolism

Yue Yang, Sergio E. Wong, Felice C. Lightstone

Biosciences and Biotechnology Division, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA, 94550, USA

Introduction

The cytochrome P450 (P450 or CYP) is a large and diverse family of membrane enzymes that catalyze the oxidation of a variety of substrates. More than 11,000 individual CYP enzymes have been identified across nearly all domains of life, including virus, bacteria, fungi, plants and animals, but were not found in *E. coli*.[1] The individual CYP enzyme is named after the encoding gene: the number immediately following the CYP abbreviation refers to the gene family, followed by a capital letter indicating the subfamily, and then a second number indicating the individual gene, e.g., CYP2E1 is expressed from gene *CYP2E1*. The CYPs oxidize not only endogenous compounds such as steroids, hormones, and other metabolic intermediates but also those exogenous or xenobiotic substrates, such as drugs or toxic chemicals. The heme-containing CYPs are essential for drug bioactivation and detoxification[2]. According to a recent survey conducted by Pfizer, CYPs account for approximately 75% of all drug metabolism[3]. Therefore, CYP catalyzed drug metabolism has drawn more and more attention. Although research focused on animal CYPs can provide helpful insight to understanding drug metabolism, the significance or relevance of such animal studies results to human tests is unclear[4]. Human CYPs are primarily expressed in the liver but are also found in lungs, kidneys and other organs[5]. Major human CYPs include CYP1A2, CYP2C9, CYP2C19, CYP2D6, CYP2E1, CYP3A4 and CYP3A5. These listed CYPs are responsible for metabolizing about 90% of drugs[5,6]. Sequence conservation across CYPs is fairly low (~20% or even lower), and only a limited number of amino acids are found absolutely conserved. In spite of this unusual variability, the general topography and structural fold of CYPs are highly conserved[7]. A highly conserved core structure is found across all CYPs, formed by a four-helix (D, E, I and L) bundle, a couple of α-helices and β-strands, and a coil[8]. This conservation might represent a common electron and proton transfer pathway, ligand binding/release groove, and oxygen activation mechanism[8].

The activity of CYPs relies on a heme cofactor (which absorbs light at 450nm, thus it is called P450, where P refers to pigment[9]). A thiolate group from a cysteine residue forms a strong electrostatic interaction to the iron at the proximal side of the heme complex, tethering it to the enzyme. The heme is bracketed between the L-helix and I-

helix, which has been recognized as the signature secondary structural feature of the CYPs[8]. The iron of the heme cofactor possesses different charge, spin and coordination states during the generic catalytic cycle operated by CYP enzymes[10]. At resting state, the CYPs are not reactive, and the heme represents a ferric-protoporphyrin complex with iron largely sitting in a low-spin doublet state[11]. At this stage, the Fe^{III} exhibits a hexa-coordinated complex, with a water molecule bound at the distal side opposite to the cysteine residue, although in some cases like CYP2D6, such a water molecule is absent[12]. The resting state heme complex is proposed to lose the water molecule upon substrate binding, forming a sextet state penta-coordinated ferric-porphyrin[13-15]. Reductase proteins could convert the resulting ferric complex into a high-spin ferrous complex, which is able to bind an oxygen molecule to form a hexa-coordinated singlet oxy-ferrous complex[14,15]. The oxy-ferrous complex is a good electron acceptor and able to go through another reduction reaction. Subsequently, the resulting ferric-peroxo anion is protonated into a ferric-hydroperoxide complex, usually called compound 0 (Cpd 0). By acquiring another proton and removing a water molecule from Cpd 0, the reactive species, compound I (Cpd I), a high valent iron-oxo radical complex, is formed and ready to react. Whether or not Cpd I is the only oxidative species is still debatable. However, a recent laser flash photolysis study on CYP2B4 and CYP119 successfully observed the Cpd I species and illustrated that Cpd I is the principle, if not the only, oxidant species in the CYP catalytic cycle[16].

The Cpd I species of CYP enzymes is responsible for the oxidization of xenobiotic compounds, such as drugs, antibiotics, or poisons. The oxidized products subsequently conjugate to water-soluble compounds, which allow them to be removed from the human body. This process is known as xenobiotic metabolism. The unsuccessful metabolism of xenobiotic compounds results in the accumulation of these compounds, which could be harmful to the human body. Many drugs are known for drug interactions – they may induce or inhibit the biosynthesis of different isozymes of CYPs, which in turn impact the activity of those CYPs. Such interactions could affect the execration of other drugs that are metabolized by the induced or inhibited CYPs and increase the adverse effects, also known as side effects. Other factors, such as the accumulation of a toxic metabolite, could also result in side effects. The side effects, especially for new drugs, are difficult to

identify until late in the clinical trial stage, leading to rejection by the Food and Drug administration (FDA) and resulting in a waste of money, time and effort. Therefore, the ability to computationally predict side effects is of great importance to drug discovery and have been the focus of many efforts in the recent years. The success of such predictive models relies on good understanding of the already known side effects. As a result, a number of efforts focused on predicting side effects have been reported[17-23].

Acetaminophen (also called Paracetamol, abbreviation APAP) metabolism by CYPs provides a good opportunity to study predictive metabolism models. As a widely used pain reliever and fever reducer, APAP is known for less side effects than either aspirin or ibuprofen in recommended dosages (1000 mg per dose or 4000 mg per day for adults)[24]. However, APAP is also well known for serious hepatotoxicity upon overdose usage, and acute usage can even lead to fatal liver damage. A number of studies were conducted to investigate the mechanism of activation and the metabolism of APAP[25-30]. While the mechanism of activation is yet not clear, the metabolism of APAP has been attributed to three pathways: (1) glucuronidation accounts for 40% to 60% of metabolism of APAP[31,32]; (2) sulfate conjugation is responsible for 20% to 40%[31,32]; and (3) hydroxylation and/or alkylation catalyzed by CYPs followed by glutathione (GSH) conjugation[33] is usually liable for less than 15% of the metabolism. Although the final products through all three pathways are inactive and nontoxic, through the third pathway CYPs can metabolize APAP into either a highly reactive toxic intermediate, *N*-acetyl-*p*-benzoquinone imine (NAPQI) or a non-toxic 3-hydroxy-acetaminophen (3-OH-APAP). NAPQI is primarily responsible for the APAP induced hepatotoxicity upon overdose usage, where detoxification pathway via GSH conjugation becomes quickly saturated. When GSH concentrations fall below 30% of its normal level, NAPQI will react and covalently bind to hepatic cells[34], resulting in the liver toxicity. Also, hepatotoxicity associated with APAP can be more severe for chronic alcoholics and/or smokers[35,36].

A number of studies on APAP metabolism by CYPs have been carried out to reveal insights of APAP metabolism by individual CYP or multiple CYPs. To date, several CYPs have been identified as capable of metabolizing APAP, including CYP1A2, CYP2A6, CYP2C9, CYP2D6, CYP2E1, and CYP3A4. Among them, CYP2E1 is widely accepted as the principle source of NAPQI[26,37], while CYP1A2 and CYP3A4 are also

believed to make significant impact to NAPQI accumulation[38]. In addition, CYP1A2 is proposed to account for the aggravation of APAP hepatotoxicity associated with smokers[39], while CYP2D6 is hypothesized to contribute to such toxicity in extensive or ultrafast metabolizers[40], though much lower activity is associated with these two CYPs compared to CYP2E1 or CYP3A4. Several of the above listed CYP enzymes could also metabolize APAP into the nontoxic 3-OH-APAP, e.g., CYP3A4 has been reported to actively produce 3-OH-APAP. Notably, different CYPs exhibit different product selectivity* on APAP metabolism: CYP2E1 turns APAP primarily into NAPQI; CYP2A6 generally modifies APAP to form 3-OH-APAP, while CYP3A4 shows comparable activity on the production of both toxic and nontoxic metabolites[41]. Due to the complicated drug interactions across CYP metabolism and the dangers of APAP hepatotoxicity, it is important to demonstrate the product preference for each CYP that metabolizes APAP. However, despite the large body of research made in this regard, the exact product ratio and/or selectivity of CYP-APAP metabolisms are still not clearly identified. Notably, experimental conclusions on APAP oxidation product selectivity for individual CYPs are either unclear, e.g., no product preferences reported for CYP3A4 that can produce both toxic and nontoxic metabolites, or confusing, e.g., CYP1A2 is labeled as producing "NAPQI (low activity) and 3-OH-APAP" in a recent review[41], whereas a ratio of 3:1 between NAPQI and 3-OH-APAP was reported for another CYP1A isozyme, CYP1A1, by Myers et al.[42].

Theoretical studies can provide useful insights to enzyme catalysis at the molecular or atomistic level. Numerous computational approaches on the ligand binding/clearance pathways, the association to the membrane[43], the electronic structures (of the heme) in the catalytic cycles[44-46], and the metabolism mechanism of CYPs[47-54], have been reported in the last decade. Such approaches are carried out at different levels of theory, including molecular mechanics (MM)[43], quantum mechanics (QM)[50,55,56], the hybrid quantum mechanics/molecular mechanics (QM/MM)[47,53,54], and coarse-grained simulations[43]. Surprisingly, there are not many molecular dynamics (MD) simulations of ligand-CYPs complexes using atomistic force fields (FFs). The lack of such MD studies might be due to the lack of high-resolution crystallographic structures of CYPs, the lack of good FF parameters for the heme system, especially the Cpd I species, and the

difficulty to derive such parameters. Recently, Cheatham and coworkers systematically developed a series of parameters for the iron-heme complexes at different stages in the CYP catalytic cycle[57]. This work opens the door for conveniently simulating drug-CYPs interaction using MD or Monte Carlo (MC) methods. In this article, we report our work on investigating the binding schemes, interaction maps and free energy profiles of APAP complexed with all CYP Cpd I species listed previously except for CYP2D6 (explained later). This systemic study of CYP-APAP complexes provides insights on APAP binding, the equilibrium distribution of different APAP binding conformations, and the product regioselectivity. Unlike most recent studies focused on one or a couple CYPs complexed with a set of ligands, our work provides a great opportunity to compare the APAP binding preferences across the set of CYPs known to metabolize APAP. In addition, the relative free energies at the binding equilibrium obtained from this work can be used toward the understanding of APAP metabolite selectivity by CYPs. Recently, Harvey and coworkers have carried out MD simulation and Density Functional Theory (DFT) calculations to investigate the selectivity of dextromethorphan by CYP2D6[58]. This study demonstrated the importance of the chemical step to the understanding of product selectivity. However, the significance of the binding schemes was less discussed. Newcomb and coworkers have found the substrate reactivities across various CYPs are largely decided by binding constants[59]. According to the Curtin-Hammett principle[60,61], a full understanding of metabolism through CYPs and selective product formation require the knowledge from not only the chemical reaction step but also the equilibrium distribution. With the experimentally measured product ratio of APAP by CYP2A6 and the computed free energy profile of the equilibrium distribution, the free energy difference between the formation of NAPQI and 3-OH-APAP is predictable. Assuming Cpd I exhibits similar inherent reactivity across different human CYP isozymes[59,62], the product regioselectivity and product ratio of APAP metabolites by various CYPs can be predicted based on the knowledge obtained from this research.

Methods and Computational Details

CYP structure selection. For better understanding of acetaminophen binding in the human metabolism cycle, only human CYP structures were selected. For each

enzyme of the following CYPs, CYP1A2, CYP2A6, CYP2C9, CYP2E1 and CYP3A4, three crystal structures with the highest resolution were selected and obtained from the RCSB protein data bank, unless less than three high-resolution structures are available. The PDB code, resolution, and complexed ligands for each selected PDB structure are listed below:

For CYP1A2, PDB ID 2HI4[63] (1.95 Å, complexed with α-naphthoflavone) was selected. This is the only high-resolution crystallographic structure available from the RCSB protein data bank.

For CYP2A6, PDB ID 2FDV[64] (1.65 Å, complexed with *N*-methyl(5-(pyridin-3-yl)furan-2-yl))methanamine), 2FDU[64] (1.85 Å, complexed with *N,N*-dimethyl(5-(pyridin-3-yl)furan-2-yl))methanamine), and 1Z10[65] (1.90 Å, complexed with coumarin) were selected.

For CYP2C9, PDB ID 1R9O[66] (2.00 Å, complexed with flurbinprofen), 1OG5[67] (2.55 Å, complexed with *S*-warfarin) and 1OG2[67] (2.60 Å, complexed with *S*-warfarin) were selected.

For CYP2E1, PDB ID 3E6I[68] (2.20 Å, complexed with indazole), 3T3Z[69] (2.35 Å, complexed with pilocarpine) and 3E4E[68] (2.60 Å, complexed with *4*-methylpyrazole) were selected.

For CYP3A4, PDB ID 3NXU[70] (2.00 Å, complexed with ritonavir), 1TQN[71] (2.05 Å, apo structure) and 3UA1[72] (2.14 Å, complexed with bromoergocryptine) were selected.

Although CYP2D6 has also been reported for APAP metabolism activity, it was not included in this study because there was no PDB structures of human CYP2D6 with good resolution (< 2.60 Å) available from the RCSB protein data bank at the time of this study. (PDB #3TBG was deposited later, in August 2012).

Structure preparation and docking APAP into CYP ternary complexes. In this step, each selected structure was processed using the protein-preparation wizard (including Prime[73,74]) in the Schrödinger suite of programs. At this stage, missing residues in the middle of the chain were added, and hydrogen atoms were assigned. An oxygen atom was manually added to the Fe–S axis at the distal side of the heme. Bond length (Fe–O), angles and orders were also manually modified to reproduce these properties reported for the Cpd I species. Subsequently, APAP was docked into each

cleaned structure using Glide[75-77] in the Schrödinger suite of programs. Ten poses with the highest ranking were saved for each structure. Further screening was made based on similarity and ranking in order to select the final five structures to be used for MD simulations. Different poses were preferred during this screening procedure. However, if all poses were already included but five was not reached, then the (next) highest ranked pose was selected, until five total poses were selected. Crystallographic water molecules were all kept, unless removed for ligand binding.

MD simulations of CYP-APAP complexes and analyses. Subsequently, all the structures kept were processed with the LeaP program in the AMBER10 suite of programs[78]. Hydrogen atoms of amino acids (AAs) and crystallographic water molecules were removed before the LeaP process. The enzymes were modeled using AMBER *ff99SB*[79], the heme cluster was modeled using the FF parameters and charge model published by Cheatham and coworkers[57], while GAFF[80] was used to model APAP. The atomic charges for APAP were computed following a procedure reported by Kollman and coworkers[81,82]. A QM optimization at B3LYP/6-31G** level was performed, followed by an electrostatic potential (ESP) calculation at HF/6-31G** level, subsequently, a two-stage restraint ESP (RESP) charge fitting procedure was conducted to derive the point charges. The resulting systems were then solvated in octahedral TIP3P[83] water boxes. Each side of the box is at least 8 Å away from the nearest solute atom. The SHAKE[84] algorithm was applied to constrain all hydrogen involved bonds, and the particle mesh Ewald (PME)[85] method was invoked to treat long-range electrostatics interactions. A weakly harmonic restraint (2 kcal/mol•$Å^2$) was applied on all enzyme heavy atoms. For each solvated ternary complex, a series of minimizations were performed in order to clear possible close contacts. Subsequently, each system was slowly heated up to 300 K over 150 ps using the Langevin thermostat[86,87]. The collision frequency of the Langevin thermostat was 5 ps^{-1}, and the time step for this stage of canonical ensemble (NVT) simulation was 0.5 fs. Followed was a constant pressure (NPT) simulation of 850 ps with a 1 fs time step to equilibrate the pressure to 1 atm. At this point, the weak restraint applied on the enzyme heavy atoms was gradually removed. Another 19 ns NPT ensemble simulation with a 2 fs time step was performed in order to well equilibrate the CYP-APAP complex. Important system properties, such as pressure, temperature and

density, were carefully examined to assure the simulation remained reasonable. A continued 20 ns MD simulation was then carried out in the NVT ensemble to propagate the system over more phase space.

For each selected PDB ID, five trajectories, excluding the initial 1 ns (150 ps NVT plus 850 ps NPT) simulations, were then merged, resulting in a 195 ns MD trajectory for each PDB ID. Subsequently, MMTSB[88] program was used to perform the cluster analysis. AMBERTools 12[89] was used to compute the RMSD and other important properties, such as key distances.

Two-dimensional (2D) umbrella sampling (USP) and free energy profiles. For each of the six CYPs, a snapshot from a set of the 19 ns NPT ensemble (with 2 fs time step) MD simulations was randomly selected as the starting structure for the following 2D USP simulations. Two reaction coordinates (RC) were selected as (1) the distance between the amide hydrogen atom (HN) of APAP and the oxygen atom (O1) coordinating to iron (RC1, $d_{\text{HN-O1}}$) and (2) the distance between the hydroxyl hydrogen atom (HO) of APAP and the O1 atom (RC2, $d_{\text{HO-O1}}$). Such a selection allows straightforward identification of conformations representing close proximity of the reactive heme complex to one of the reacting centers from APAP, either the amide nitrogen or the 3-carbon adjacent to the hydroxyl group. In fact, such a close proximity of reactive atoms/groups should usually correlate to the product preferences or regioselectivities.

For each CYP, sampling windows were placed along RC1 from 1.6 Å to 11.8 Å and RC2 from 1.5 Å to 12.3 Å with a 0.6 Å interval for both RCs. Regions that general MD simulations never reached were excluded. As a result, a total of ~280 windows were prepared for each CYP isozymes. A NPT ensemble simulation of 10 ns was carried out at each window in order to investigate the Gibbs free energy surface. A 20 kcal/mol•Å^2 harmonic potential was applied on each RC for all the simulations, and data were collected every 100 steps (200 fs) during the last 5 ns MD at each window. The weighted histogram analysis method (WHAM)[90-92] was conducted using Grossfield's WHAM code[93] to remove the biased potential and reconstructed the free energy profiles. Statistical computing and graphics program R[94] was used to generate the plots for each

profile. In order to analyze the CYP-APAP interaction pattern, a short 2ns unbiased MD simulation in the NPT ensemble was carried out for each important binding scheme.

A graphical illustration of the entire procedure described above can be found in Scheme 1.

Results and Discussions

APAP in CYP3A4 features two stable binding states connected by a less stable intermediate state and a 'flat' free energy surface. Among all the isozymes, CYP3A4 draws most attention because it can metabolize an assortment of substances, including a number of large substrates. Crystallographic studies reveal that CYP3A4 has a voluminous active site that can bind large ligands. The big binding pocket also allows CYP3A4 to bind multiple identical ligands or different ligands simultaneously, e.g., two APAPs or APAP and caffeine[95]. The ability to bind multiple ligands simultaneously makes CYP3A4 a good target to study some drug-drug interactions, e.g., the impact of caffeine on APAP metabolism. Such investigations are currently underway in our lab and thus will not be discussed in this article.

Docking ligands, especially small ligands, against CYP3A4 could be difficult due to the large size of the pocket. Not surprisingly, APAP exhibits great flexibility in CYP3A4: it adopts more than 5 different orientations in the 10 highest-ranking poses for 3NXU, 1TQN and 3UA1 separately. However, as mentioned previously, only the five highest-ranking poses were kept for the subsequent studies. Similarities were observed across three sets, as at least one pose featured HN closely pointing at O1 and one pose with HO in close proximity to O1 were identified for each set (Figure 1a). Differences were seen across three sets, which is not surprising because of the slightly different binding pocket structures, making our choice of using 3 PDBs more rational. Different PDB structures allow us to examine more possible APAP binding poses, thus making our results less dependent on the choice of starting structure. MD simulations were then carried out to investigate the dynamics of APAP in CYP3A4. Several AA side chains surround and interact with APAP in the binding sites, including R105, F108, S119, F213, F304, A370 and L482. However, none of these side chains are in close contact with

APAP, resulting in a rather spacious pocket in which APAP can likely rotate to change orientations.

By visually checking the trajectories: several major conformations of APAP binding are easily identified: (1) HN in close proximity to O1, (2) HO in close proximity to O1, (3) APAP still close to O1 but neither HN or HO are directly in close proximity to O1 (Figure 1b). Cluster analysis further confirmed these three binding poses. Distribution of RC1 and RC2 was extracted and plotted (Figure 1c). Clearly, each one of the poses identified from docking and MD analysis has a fairly significant distribution. Among them, three clusters have the largest distributions. We termed these three clusters as S1, S2 and S3 and labeled them in Figure 1c (the same terminology will also be applied to other CYP-APAP systems to represent similar conformations). S1 shows a shorter RC1 but longer RC2, corresponding to the binding pose (Figure 1b, left) that features a close proximity between atom HN and O1. S2 exhibits a shorter RC2 but long RC1, in accord to the binding conformation (Figure 1b, middle) that features a close proximity between atom HO and O1. S3 (Figure 1b, right) seems to function like an intermediate state connecting S1 and S2. Remarkably, the 'traffic' between S2 and S3 seems heavier than that between S1 and S3, suggesting possible faster interconversion. This interesting observation implies that the energy barrier between S2 and S3 is lower than that between S1 and S3.

Although there are 15 sets of 40 ns MD simulations, starting from different binding poses docked against different crystal structures and giving a great deal of useful insights into various APAP binding states, the MD simulations still cannot provide sufficient information to help us fully understand the APAP binding in CYP3A4. A good example is the interconversion observed between S3 and two more populated states, which qualitatively illustrate the connections between those binding states. However, in order to quantitatively understand the underlining kinetics, more robust free energy studies are required. More importantly, we found that each MD trajectory is fairly 'biased' by its starting structure. By plotting the distance distribution of the first 5 ns of simulation for each of 5 MD simulations for a single PDB in a different color (Figure 1d), each of the MD trajectories clearly samples only a limited amount of phase space close to the initial structure. Hence, more starting points would be recommended for studying enzyme

dynamics using MD simulations, but even starting from tens of initial structures still could not guarantee sampling the entire energy landscape. As a consequence, the QM/MM or QM calculation, starting from snapshots taken from these MD trajectories, possibly only represents a local structure. Therefore, using more robust free energy simulations to thoroughly study the thermodynamics of protein-ligand binding not only provides valuable insights into the individual interactions between ligand and the environment, but also identifies reliable starting points for more expensive QM/MM or QM studies to investigate the chemical reaction mechanism.

With this motivation in mind, subsequent USP simulations were carried out to further investigate the diverse APAP binding conformations and discover the underlying free energy relationship. The sampling windows were placed based on the distribution of two RCs, and the range of each RC was selected to be able to cover the APAP movement in the entire binding pocket (Figure 2a). The WHAM analysis was performed to unbias the RCs distribution and reconstructed the free energy surface. The resulting free energy profile is plotted in Figure 2b. Clearly, three minima are identified and their locations are consistent with S1 (RC1 = ~2 Å RC2 = ~8 Å), S2 (RC1 = ~8 Å, RC2 = ~2 Å) and S3 (RC1 = ~6 Å, RC2 = ~7Å), as classified from the RC distribution map. The S1 state has the lowest free energy, corresponding to the most stable conformation. The free energies associated with the S2 and the S3 state are approximately 0.5 and 1.3 kcal/mol higher than that of the S1 state, even though S2 seems to have a larger density than S1 in the previously described distance distribution map (Figure 1c), suggesting it should have the lower free energy. Based on this profile, the barriers on the S1 <-> S3 and S2 <-> S3 pathways are almost identical in absolute energy, but the fact that S2 has a higher energy than S1 makes the barrier height of S2 <-> S3 about 0.5 kcal/mol smaller than S1 <-> S3, meaning the interconversion between S2 and S3 is indeed faster than that between S1 and S3. The S2 to S3 transition only requires a hydrogen bond (H-bond) breaking between the hydroxyl and heme oxygen and a small rotation (Figure 3, see also Figure S1 in Supporting Information). In fact, the ~0.7 kcal/mol barrier height identified for S2 to S3 conversion is consistent with the energy associated with the loss of a H-bond (~1 kcal/mol). Notably, the entire free energy profile is fairly flat and contains several

minima-like basins. Such a fact is not completely surprising, because CYP3A4 is known for its voluminous active site and promiscuity for ligand binding.

The structures of the CYP3A4-APAP complex at these three states are shown in Figure 3. At the S1 state (Figure 3a), the aromatic group lays in a plane vertical to the heme group while having the amide nitrogen in close proximity to the reactive oxygen. Unlike in both docking and MD stages where APAP looks extended, at S1 state identified from USP simulations, the amide group of APAP is bended away from the phenol plain. This APAP configuration nearly reproduces the pose that the Nelson lab proposed based on their NMR T_1 paramagnetic relaxation experiment[95], expect for the heme propionic acid groups pointing toward different directions. In fact, only the distances of the aromatic hydrogen atoms and the methyl group to the iron center were measured in their experiments. Thus, the orientation of the carboxylic groups might not necessarily represent the most realistic one. The S3 state (Figure 3c) is directly connected to both the S1 and S2 state with fairly low barriers, while the barrier separating the S1 and S2 state (Figure 3b) is much higher, indicating it is indeed functioning as an intermediate state. However, at this state, one of the two possible aromatic hydroxylation centers (two aromatic carbon atoms ortho to the hydroxyl group) is still in close proximity to the heme oxygen, making S3 also a possible candidate for 3-hydroxylation. It is also worth noting that RC2 almost keeps fixed during S1 to S3 transition while RC1 only slightly changes during S3 to S2 transition. Such an observation can again be related to the large degrees of freedom ligands have in the active site of CYP3A4, as flexible rotation is allowed for APAP. In fact, only a few AA side chains are within the 4 Å range to APAP during these transitions, namely R105, S119, I369, A370 and L482; thus, the steric effect felt by APAP should be limited. At S2 state, APAP nearly stands orthogonal to the heme group with the hydroxyl group pointing toward the reactive oxygen. The methyl group is surrounded by several phenylalanine residues, namely F108, F213, and F304 that belong to the unique phenylalanine cluster[96]. Though the exact function of these Phe residues to the APAP methyl group is not yet clear, F108 and F213 appear to help stabilize the APAP orientation through weak CH-π interactions[97] in this specific binding state. Once again, only a few other AA side chains are found present in close range to APAP, giving the ligand plenty of flexibility to move around. Such loose contacts between the protein

and ligand possibly explain why CYP3A4 is not the major CYP responsible for APAP metabolism, despite being the most abundant CYP in human body, as the ligand could not gain high affinity through non-bonded interactions with the enzyme environment. According to the relative free energy, the concentration ratio of these three conformations (S1:S2:S3) is close to 6:2:1. Therefore, unless the barrier for *N*-oxidation or 3-*C*-hydroxylation is much higher than the other, notable activity for both metabolites would be expected. This conclusion is consistent with the fact that CYP3A4 is not a primary NAPQI producer.

Individual interactions between ligand and protein side chains are valuable in studying the protein-ligand binding. The free energy simulations are capable of providing such important insights. In general, R105 and S119 are localized at the propionic acid side, and T309, I369, and A370 stand like a wall on the other side, encircling the APAP pocket, while F304 and L482 appear to form the ceiling for the pocket (Figure 3). In all three states, the phenol group is heading toward the hydrophobic wall formed by T309, I369, and A370, and L482 is in close contact to the phenol group of APAP. At both the S1 and S3 state, APAP extends itself in an orientation parallel to the heme plane with the phenol group vertical to that plane. With this orientation, the methyl group is in close contact to residues R105 and S119. In contrast, at the S2 state APAP exhibits a vertical orientation so the methyl group is no longer close to R105 and S119 but rather interacts with F108 and F213. The relative position of the methyl group to the two phenylalanine side chains falls into the range of CH-π interactions, which could help stabilize this configuration (Figure 3b). F304, another Phe located near APAP, is close to the aromatic group of the ligand and possibly interacts with the ligand via steric effects. This aromatic side chain is not close to APAP in either of the two other binding states. However, in S1 state a water molecule is found to mediate interesting interactions between APAP and the protein environment. In such a scheme, APAP forms a H-bond to a water molecule with the hydroxyl group as the H-bond donor, and one of the water hydrogen atoms forms a H-bond to the backbone of G481 while the other hydrogen points to the aromatic ring center of F213 (Figure 3a). These interactions might further stabilize the configuration, in addition to having a direct interaction between the APAP amide group and the heme oxygen, making the S1 state the most stable. This water-mediated interaction system and

the amide-oxygen interaction are absent from the S3 state (Figure 3c), decreasing its stability.

APAP exhibits a remarkably dominant S1 state in CYP2E1. Even though CYP2E1 is a less abundantly expressed CYP in human body compared to CYP3A4, human CYP2E1 is responsible for 30% to 78% of APAP metabolism, with most of the product being the toxic NAPQI[26]. Despite the number of studies that show its critical role in the hepatotoxicity associated with APAP metabolism, the high specificity and product selectivity of CYP2E1 for producing NAPQI are not yet clear.

In order to obtain the atomistic level of detail for molecular recognition of APAP by CYP2E1, we carried out the same workflow described previously to study the CYP3A4-APAP complex. Surprisingly, the docking results did not give more S1-like poses (with a short RC1) than S2-like poses (with a short RC2), despite the enzyme being well known for metabolizing this ligand to its toxic metabolite. Even more so, the top-scoring pose is S2-like, which should lead to the non-toxic metabolite (Figure 4). This result further illustrates the inappropriateness of using docking methods alone to study the CYP-drug binding and interactions. However, through analysis of MD trajectories, we clearly see more population of the S1 like conformation than the S2 like configuration. In addition, a S3 like cluster is not observed at all, which implies that fast interconversion between different states is probably prohibited in this complex. We identified more AA side chains within 4 Å to the ligand than in the CYP3A4-APAP complex via a visual check of the active site snapshots from MD trajectories (Figure S2). Hence, we speculate that a much tighter binding pocket is formed in CYP2E1 than in CYP3A4. The transition between these two conformations is likely restricted by the steric strain imposed by such a compact pocket. The free energy profile (Figure 5a) of APAP in CYP2E1 confirms the dominance of the S1 state, as well as, the absence of intermediate state S3. The NAPQI precursor (S1 state, RC1 = ~2 Å, RC2 = ~7 Å) is approximately 4.2 kcal/mol more favorable in free energy than its non-toxic counterpart, 3-OH-APAP (S2 state, RC1 = ~7Å, RC2 = ~2Å). The absence of the intermediate state results in a large free energy penalty for the transition between two conformations, at about 7.3 kcal/mol. With such an energetic cost, the barrier crossing activity is expected to be ~1000 times slower than similar transitions taking place in CYP3A4. Notably, the free energy increases much

more rapidly upon the ligand moving further away from the binding pocket than seen in the case of CYP3A4, implying that APAP has higher off binding rate in CYP2E1 than CYP3A4. Given the fact CYP2E1 is expressed much less than CYP3A4 but plays a more essential role in APAP metabolism, high affinity between this ligand and CYP2E1 is a likely explanation.

Taking a closer look at the active site structure at the two different binding states, the S1 state (Figure 6a), with the methyl group heading 'down' toward the heme while the phenol group heading 'up' to the outside of the pocket, is different compared to the S1 state in CYP3A4, where the phenol group is closer to the heme than the methyl group. Also, at the CYP2E1 S1 state, APAP lays more orthogonal to the heme plane than that in CYP3A4. Most interestingly, the APAP aromatic ring takes the same physical space in the CYP2E1 binding pocket for each of the two binding states, S1 and S2 (Figure 6b), even though the remaining part of molecule is in opposite directions (Figure S2). Again, the CYP2E1 binding site is so compact that APAP is strongly restrained in a narrow groove, providing very specific interactions. In a recent binding study where long chain fatty acids binding into CYP2E1 were experimentally determined[98], the ligands were found to bind in a similar groove as what we observe in this study; and more interestingly, that groove can be extended further out to accommodate long-chain ligands, which might be the small drug binding/unbinding pathway. Moreover, as observed in unbiased MD simulations, we also noticed that more AA side chains are in close range to the drug molecule than in the CYP3A4 case. In detail, I115, F116, F207, L210, D295, F298, A299, T303, L363, V364, and L368 form a tight site, leaving only a narrow groove for APAP. In addition, not surprisingly, nearly identical groups of AA residues are interacting with APAP at the two states, since the ligand occupies the same 3-D space. At the S1 state, an interesting water mediated H-bond system is found to stabilize the ligand carbonyl group and L363, and a multiple water involved H-bond system is also identified to build a connection between the APAP hydroxyl group to H109, N206, and D295 (Figure 6a). A294 can also be included into this water involved H-bond pattern (Figure S2), though four water molecules separate its backbone carbonyl oxygen from the APAP hydroxyl group. At the S2 state, a little different two water mediated H-bond system is seen to stabilize the amide group of APAP to D295, while the indirect interaction between L363

and APAP is missing. Also, the methyl carbon atom of APAP is placed only 3.5 to 4 Å away from several heavy atoms of F207, L210 and F298, potentially producing strong steric repulsion. The relative position of the methyl or other group of APAP to the two phenylalanine residues are not favorable for any π-interactions. As a result, the S2 state is expected to be much less stable than the toxic precursor S1 state.

Overall, the MD simulations and the free energy profile generated from the USP are able to illustrate the difference in the binding environments that APAP encounters between the active binding sites of CYP2E1 and CYP3A4. Also, the relative free energy difference between the different binding states and the free energy changes upon pulling the ligand away from the binding site and clearly explains why CYP2E1 has a higher specificity and more particular product regioselectivity than CYP3A4. Additionally, docking methods alone, although straightforward and computationally inexpensive, alone cannot provide adequate insight in understanding drug metabolism by CYP.

APAP displays an opposite (to CYP2E1) regioselectivity in CYP2A6. CYP2A6 is well known as the primary CYP enzyme that metabolizes nicotine oxidation and is involved in the metabolism of APAP as having a less important or minor role[37]. The preferred metabolite of APAP produced by CYP2A6 is the nontoxic 3-OH-APAP, though the toxic metabolic NAPQI is also generated by this enzyme at approximately a 1:3 ratio (to the major product)[99]. This selectivity can be explained by the free energy profile generated from our USP simulations (Figure 5b). Clearly, the S2 state (RC1 = ~7 Å, RC2 = ~2 Å) has the lowest free energy in the profile, while the location where the S1 state should be (RC1 = ~2 Å, RC2 = ~7 Å) does not represent a stable energy basin in CYP2A6. A unique free energy basin is identified at position RC1 = ~ 4 Å, RC2 = ~4 Å. This new basin possibly functions as a replacement for the S1 state (S1r), which orientates APAP ready for the toxic *N*-oxidation. The free energy associated with the S1r state is about 2.5 kcal/mol higher than the S2 state. Another basin, whose function is possibly similar to the S3 state found for CYP3A4, is located at RC1 = ~8 Å, RC2 = ~4 Å. This S3 state is about 1.5 kcal/mol less stable than the S2 state, providing an energetically favorable pathway for the S1r <–> S2 transition. The protein-ligand, and ligand-solvent interactions at these binding states (Figure 7) can help us understand this unique free energy landscape exhibited for CYP2A6. At the S1 state (Figure 7a), the

APAP methyl group is strongly pushed by F118, I300, I366, and L370 to a position that causes steric repulsion to the APAP phenol group, while at the S1r state (Figure 7b), this repulsion is relieved with the methyl group moving away from the phenol group. The distances between the methyl carbon atom and the closest aromatic carbon atom are 2.7 and 3.1 Å at the S1 and S1r state, respectively. In addition, at the S1 state, the amide group forms a water-mediated H-bond network with the backbone carbonyl oxygen of N297, while at the S1r state, the carbonyl oxygen forms a direct H-bond to the side chain of the same AA residue. Furthermore, a water-mediated H-bond network that stabilizes the hydroxyl group to the backbone of I366 is observed at the S1r state, but such a H-bond pattern has not been found in S1. Considering the role that I366 played in the missing H-bond and repulsing the methyl group into a sterically unfavorable position, the S1r state is much more stable than the S1 state. APAP is further relaxed from those steric repulsions at the S2 (Figure 7c) and S3 states (Figure 7d), resulting in lower free energies. N297 forms a direct H-bond and a water-mediated H-bond network with the APAP carbonyl oxygen in the S3 and S2 state, respectively, helping stabilize the ligand. In addition, a direct H-bond stabilizing the ligand hydroxyl group is also observed in the S2 and S3 states, with the acceptor being the heme oxygen and T305 side chain, respectively. A direct H-bond from APAP to the reactive oxygen can tightly tether the ligand to the heme center, giving a closer proximity between the reactive oxygen and the ligand reaction center. Moreover, in the S2 state, F209 is perfectly placed such that a weak CH-π interaction can form with the ligand methyl group. Although several protein side chains, namely F107, F111, F116 and I300, can also form CH-π interactions with APAP, the relative position between the aromatic group and the corresponding CH groups is not as well placed as F209 in the S2 state. As a result, the S2 state is the most stable binding state in the CYP2A6-APAP complex, where S3 is more energetically favorable than the S1r state.

Because the APAP metabolite ratio by CYP2A6 has already been experimentally measured[99], the free energy profile of CYP2A6-APAP binding complex provides a great opportunity to understand the CYP-APAP metabolism mechanism in terms of the free energies of the two reactive conformations at the binding stage and in the chemical step. Although such a product ratio was also measured for CYP2E1 in the same work, the ratio

for CYP2E1 is strongly impacted by the presence of cytochrome b_5, while the ratio for CYP2A6 metabolism is almost unperturbed[99]. Therefore, we believe CYP2A6 is more appropriate than CYP2E1 to be used to understand the free energy landscape for the metabolism. According to the Curtin-Hammett principle[100], the product distribution of APAP metabolism is determined by both the equilibrium distribution at the binding state and the free energy difference at the chemical step. According to the 1:3 product ratio identified experimentally[99], the free energy difference in the metabolism cycle, including both the binding and the chemical step, should be about 0.8 kcal/mol favored for the non-toxic metabolism pathway. Thus, based on our calculations, we predict that the reaction barrier for the *N*-oxidation would be lower than the 3-hydroxylation with an approximately 1.7 kcal/mol in free energy difference, as shown in Scheme 2. In fact, the aromatic hydroxylation would be expected to have a higher reaction barrier than its counterpart reaction, since in the aromatic hydroxylation, the C-O bond formation would unavoidably impact the hydroxyl to iron oxygen interaction, leaving the ligand less stabilized. Overall, although the chemical step energy barrier favors the *N*-oxidation over the aromatic hydroxylation, the 2.5 kcal/mol free energy difference between the two binding equilibriums still make the aromatic hydroxylation the favorable pathway, resulting 3-OH-APAP preference. Such an estimated reaction barrier height difference could be applied to understand the product selectivities by different CYPs (since similar heme reactivity is proposed for different CYPs), although a systematic investigation of the chemical steps is necessary to more accurately predict the product ratio.

An interesting remote site was observed on the CYP1A2-APAP binding profile. As previously mentioned, the product selectivity of APAP metabolism by CYP1A2 is difficult to conclude from the currently available experimental results. Monostory and coworkers reported that the inhibition of CYP1A2 and CYP3A4 in vitro failed to reduce NAPQI production[101]. On the other hand, Zaher et al. found CYP1A2 and CYP2E1 double-null mice were protected against APAP caused toxicity[102]. Snawder et al. concluded that CYP1A2 activity increased at high dose of APAP[30], and in contrast, Wright and coworkers claimed that CYP1A2 inducers did not increase the NAPQI production after therapeutic dosage of APAP among human volunteers[4]. These claims

look confusing and a bit contradictory. As a result, the CYP1A2-APAP binding preference and product selectivity remain unclear.

We constructed the free energy profile (Figure 5c) for APAP binding in CYP1A2 following the same strategy described previously. Based on this profile, the S2 state (RC1 = ~7 Å, RC2 = ~2 Å) is more than 2 kcal/mol more favorable than the S1 state (RC2 = ~2 Å, RC2 = ~7 Å). The two states are connected by an intermediate state, S3 (RC1 = ~9 Å, RC2 = ~8 Å), at approximately 4 kcal/mol higher than the S2 state. The interconversion between S3 and each productive conformation needs to overcome a ~5 kcal/mol (relative to the S2 state) transition barrier. The direct interconversion between the S1 and S2 states is possible but at a higher free energy cost (~6 kcal/mol). Although the CYP1A2 S3 and CYP3A4 S3 states both function as intermediate states, the S3 state in CYP1A2 is notably much further away from the heme center than the CYP3A4 S3 state. This is not entirely unexpected because the relatively small binding pocket of CYP1A2 prohibits the flexible rotation of APAP. An interesting observation of this free energy profile is the identification of a distal site (SD) at RC1 = ~5 Å, RC2 = ~11 Å. Such a distal site has been reported in other enzymes, such as human carbonic anhydrase II (HCA2)[103] and could correspond to an intermediate state along the binding channel, though the exact function of this distal site is not yet clear. Our free energy map points out that the non-toxic conformation, S2, is the energetically favorable binding state in the CYP1A2-APAP complex. Therefore, unless the reaction barrier for aromatic hydroxylation is more than 2 kcal/mol greater than the reaction barrier for N-oxidation (which is unlikely based on the results from CYP2A6), we should expect CYP1A2 to exhibit 3-OH-APAP preference over NAPQI, which agrees with the conclusion that Rendic listed in his review[41].

Similar to CYP2E1, CYP1A2 is also known for a compact binding site and a narrow entry channel. Hence, the APAP binding locations at all four binding states fall within a narrow groove (Figure S3). In addition, steep free energy change upon moving APAP away from the heme cofactor is again observed, similar to that seen in CYP2E1, suggesting the ligand gains high affinity upon binding and needs to overcome a high energy barrier for off binding. Consistently, CYP1A2 is also found to be a major metabolizer of APAP, second only to CYP2E1[26]. S3 is distant from the heme center, illustrating flexible rotation of APAP and fast interconversion between S1 and S2 are

likely prohibited in this enzyme, as well as in CYP2E1. The locations of the SD and S3 states are close, implying such a site indeed can function as an intermediate state during the ligand binding process or ligand orientation conversion. The fact that SD and S1 are rarely overlapping makes it possible to hypothesize CYP1A2 could simultaneously bind two APAPs, or APAP along with an other small ligand, such as nicotine. Although this assumption of multiple APAP binding seems able to help explain the observation that CYP1A2 activity (not limited to the NAPQI production) increases at higher dose of APAP, the removal pathway of the metabolite seems blocked upon ligand binding at SD location.

At the S1 state (Figure 8a), the APAP phenol group is flanked by several hydrophobic residues, including L382, T385, I386, L497, and T498, while the methyl group is surrounded by F226, A317, and D320. A direct H-bond between the ligand amide group and the heme oxygen helps tether the ligand to the heme center. At the S2 state (Figure 8b), side chains of A317, T321, L382, I386, L497, and T498 encircle the ligand phenol group, and T223, F226, and V227 are within close range to the methyl group. In addition to a H-bond between the APAP hydroxyl group and the heme oxygen, another direct H-bond between the backbone of G316 and the amide group can help APAP gain more stability. At state S3 (Figure 8c) or SD (Figure 8d), the ligand is fully surrounded by enzyme side chains, including F125, T223, F226, V227, G316, A317, D320, L497, and T498. A couple of direct H-bond interactions, namely APAP hydroxyl to D320 carboxyl and APAP amide to G316 backbone, help stabilize the S3 state. However, at the S3 state APAP is compactly flanked by G316, A317 and L497, and the possible steric repulsion between the hydrophobic side chains and the ligand can decrease the stability of this intermediate state, making it energetically less stable than the two reactive states. At the SD state, the methyl group of APAP is sandwiched by the aromatic rings of F226 and F260, potentially favored for CH-π interactions. In addition, the ligand is further stabilized via water-mediated H-bond patterns between its amide oxygen and AA residues including T118, N312, and D313. The ligand is relaxed in such a secondary pocket, as no clear steric repulsions can be found.

Non-toxic metabolism preference, a flat free energy surface and a bigger and more stable distal site (than seen in CYP1A2) are found in the CYP2C9-APAP

complex. CYP2C is another big human CYP subfamily, accounting for approximately 18% of hepatic CYP content[104,105]. Within this subfamily, CYP2C9 is the principle isozyme expressed in the human liver[104]. CYP2C9 does not seem to play an important role in APAP metabolism, as Patten et al. claimed no APAP activation was observed in their study across a set of human CYPs[27]. However, it is worth to mention that in the same study, CYP2D6 was also found to have no activity[27], although a later study by Nelson and coworkers confirmed CYP2D6 is involved in APAP bioactivation[40]. Similarly, a recent review listed CYP2C9 as able to metabolize APAP to NAPQI at low activity and 3-OH-APAP at low rate[41].

With only docking results and MD simulations, the binding selectivity of the CYP2C9-APAP complex is inconclusive because conformations that lead to both NAPQI and 3-OH-APAP are observed. In fact, CYP2C9 is also known for having a large active site. APAP is expected to have many degrees of freedom and show multiple binding conformations. Similar to CYP3A4, which also possesses a big binding pocket, the free energy profile (Figure 5d) reveals CYP2C9-APAP binding complex is having a fairly flat free energy surface. Both the S1 (RC1 =~ 7Å, RC2 =~ 2Å) and S2 (RC1 =~ 2Å, RC2 =~ 7Å) states can be easily located on this profile. A S3 state can also be found at RC1 =~ 9Å, RC2 =~ 4Å. However, unlike the S3 state in CYP3A4, the CYP2C9 S3 state seems no longer an intermediate state for the interconversion between the S1 and S2 states because the S1 <-> S2 conversion routes that does not pass S3 is identified and is energetically favorable. Instead, the S3 state in the CYP2C9-APAP complex is more likely serving as a local intermediate along the ligand binding pathway. Energetically, the S2 state has the lowest free energy, meaning this enzyme prefers the non-toxic binding scheme. The free energies of the S1 and S3 states are approximately 2.0 and 1.6 kcal/mol higher than the S2 state. Considering the free energy difference for the chemical step of *N*-oxidation and aromatic hydroxylation is less than 2 kcal/mol more favorable for the toxic metabolite (as concluded from the previous CYP2A6 study), both metabolites should exhibit comparable activity, although the NAPQI formation might be a little bit slower. In addition, similar to CYP3A4, the flat free energy surface implies APAP may not gain high affinity upon binding; thus 'low rate' and 'low activity', as listed by Rendic[41], are explained. A distal site (SD) is also observed in CYP2C9 and, according to

the energy landscape, it is much bigger than that seen in CYP1A2. The lowest free energy basin at SD is only less than 1.0 kcal/mol higher than S2, making it possible to serve as an intermediate on the binding pathway or a reservoir for ligand binding, similar to the function proposed for the low affinity CO_2 binding site in HCA2[103]. In fact, CYP2C9 is reported to being able to bind multiple ligands simultaneously[106].

Within the heme binding pocket in CYP2C9, APAP is surrounded by V113, I205, D293, G296, A297, E300, T301, L362, L366, and F476. At the S1 state (Figure 9a), all the side chains listed above are in close contact with the ligand. The amide group is tethered to the heme through a H-bond to the reactive oxygen. In addition, water-mediated H-bond patterns are found to stabilize the APAP hydroxyl group and carbonyl oxygen to the backbone of E300 and the carboxyl group of D293, respectively. The ligand is further stabilized via a two water involved H-bond system that connects the ligand carbonyl oxygen to the side chain of R108. At the S2 state (Figure 9b), L366 and F476 move to 4 Å away from the ligand, possibly resulting in less steric repulsions. The APAP hydroxyl group forms a direct H-bond to the heme oxygen, tethering the ligand to the heme cofactor. Another direct H-bond between the APAP amide group and the backbone oxygen of G296 further stabilizes the ligand. At the S3 state (Figure 9c), APAP is found to be surrounded by residues R108, V113, I205, V237, M240, V292, D293, G296, A297, L366, and F476. Neither the APAP hydroxyl group nor the amide group directly interacts with the reactive oxygen. Instead, a water-mediate H-bond is found to bind the ligand hydroxyl group to the heme cofactor. In addition, the ligand amide group and the backbone oxygen of V292 might form a direct H-bond, although the AA oxygen and amide nitrogen are approximately 3 Å away. It is worth noting that APAP has many degrees of freedom in the CYP2C9 binding pocket. Almost the exact same residues surround the ligand at the S1 and S2 states, making it rational to assume that the interconversion between the S1 and S2 states does not require APAP to move far from the heme binding pocket. Therefore, the S3 state can further be confirmed as a binding intermediate instead of functioning as an intermediate state for the S1 <-> S2 conversion, as in CYP3A4. In fact, the locations of the S3 and SD states (Figure 9d) are close, making the above assumption more reasonable. At the SD state, APAP is further away from the heme binding pocket and surrounded by enzyme residues, including F100, L102,

R108, V133, N204, I205, L208, V292, D293, and F476. The ligand is interestingly positioned such that the C_δ atom of the I205 side chain possibly forms a weak CH-π interaction with the aromatic ring of APAP, and the hydroxyl group of APAP is pointing to the aromatic ring of F476. Also, the carbonyl oxygen forms a direct H-bond to the side chain of N204. The ligand might be able to gain affinity through these interactions, resulting in this SD state suitable for an intermediate state along the binding route. The vacancy of the ligand is filled out by a structured water system because four water molecules are found to form a H-bond relay that connects the heme oxygen and the backbone carbonyl oxygen of I205.

The CYP2C9-APAP complex exhibits a free energy surface even flatter than that seen in the CYP3A4-APAP system because the free energy change upon moving the ligand toward the entry of the pocket is less than 3.0 kcal/mol. Therefore, it is rational to assume APAP gains less affinity upon binding in CYP2C9 then in CYP3A4, explaining why APAP metabolism by CYP2C9 is identified experimentally as 'low activity' and 'low rate'. The non-toxic conformation is energetically more favored (~3 kcal/mol) than its toxic counterpart. Even though the toxicity-leading *N*-oxidation has a lower barrier (< 2 kcal/mol) than the non-toxic aromatic hydroxylation, CYP2C9 is still expected to exhibit preference for 3-OH-APAP over NAPQI.

Summary. Individual CYPs have different site-of-metabolism selectivities on APAP metabolism. However, the source of such different binding preferences is not yet clear. In this work, we have shown that each CYP preferentially binds different APAP conformations that are consistent with different sites of metabolism. Combining the relative binding free energies for each CYP with the chemical reaction barrier difference (obtained from the CYP2A6 case study) can improve the understanding of APAP metabolite selectivity by CYPs. By carefully analyzing the active site structures and the protein-ligand interactions at each binding state across a set of five CYPs, we conclude that the shape of the cavity and the individual AA side chain to ligand interactions, and deciding factors in the CYP selectivity.

To achieve the best affinity and selectivity, the ligand conformation should be able to complement the size and shape of the binding pocket. For CYPs that possess voluminous active sites, such as CYP3A4 and CYP2C9, the relatively small ligand APAP

has many degrees of freedom, especially in the CYP3A4 pocket where fast interconversion via flexible rotation of APAP between S1, S3 and S2 states is allowed. The energy difference between two productive binding conformations is less than 2 kcal/mol. The ligand feels few direct interactions from CYP3A4 and CYP2C9, consequently affecting the binding affinity. While in CYP2E1 and CYP1A2, with the well-known compact binding sites, a steeper free energy change from the edge to the center of binding pocket is observed, suggesting with more contact with the protein, APAP might be able to gain more affinity in these two cases than in CYP3A4 or CYP2C9. Also, the interconversion between the two reactive binding states is either disallowed in CYP2E1 or only allowed at a much longer distance away from the heme center. Feeling more protein-ligand interactions forces APAP to adopt a single preferred binding pose compared to others, usually by more than 2 kcal/mol. In CYP2E1 the toxicity-leading S1 state is preferred, while in CYP1A2 and CYP2A6, the S2 binding state is more favored.

For the same APAP binding states across different CYPs, comparing the distances between according reactive center to the heme oxygen gives interesting correlation to the product preference. A good example is the S1 state in CYP2E1-APAP complex. At this state, the amide hydrogen is perfectly orientated to the heme oxygen such that a H-bond can be formed to tether the ligand amide group to the heme center. The distance between amide nitrogen and heme oxygen is only 3.2 Å, approximately 0.2 Å shorter than the same distance in CYP1A2 and CYP2C9, while more than 0.7 Å shorter than that in CYP2A6 and CYP3A4. Not surprisingly, CYP2E1 is the most dominant contributor to the toxic NAPQI accumulation. On the other hand, at the S2 state across different CYPs, the distances between hydroxyl hydrogen and the heme oxygen are much more similar, with CYP1A2, CYP2A6 and CYP2C9 all at approximately 2.0 Å, while CYP2E1, who principally converts APAP into NAPQI, and CYP3A4, who is not a major 3-OH-APAP producer, at approximately 2.2 Å. Although CYP1A2, CYP2A6 and CYP2C9 all giving comparable distance between the APAP hydroxyl hydrogen to the heme oxygen, they are showing quite different distances between the actual reaction centers, the C_3 atom on the phenol ring and the heme oxygen. The shortest distance of C_3-O belongs to CYP2A6 (3.3

Å), who is well known as the principle 3-OH-APAP generator, with CYP2C9 a bit longer (less than 3.4 Å) and CYP1A2 the longest (~3.5 Å).

Although the structurally conserved CYPs show little sequence conservation, comparing the protein-APAP interactions across different CYPs still draws our interest and, in fact, provides some valuable insights. First, all CYPs listed in this study invoke interactions with aromatic side chains, mostly phenylalanine, e.g., F108, F213, and F304 in CYP3A4; F116, F207, and F298 in CYP2E1; F107, F111, F116, and F209 in CYP2A6; F125 and F226 in CYP1A2; and F100 and F476 in CYP2C9. In addition, a unique Phe-cluster distantly composed of 7 Phe residues above the heme pocket has been identified for the CYP3A4[96], and a similar but much smaller cluster containing 3 Phe residues (F100, F114 and F476) is observed for CYP2C9[107]. Although the exact function of the Phe-cluster and the Phe residues close to the APAP binding pocket is not conclusive without a specific study, an educated and aggressive assumption is those aromatic side chains can steer APAP orientation during the course of binding to encourage the best fit to the pocket. Second, Leu and Ile residues located near the K helix, one of a limited number of sequence conserved regions, are found to play an important role in APAP binding. Examples include L369 in CYP3A4, L363 and L368 in CYP2E1, I366 and L370 in CYP2A6, L382 and I386 in CYP1A2, and L362 and L366 in CYP2C9. APAP can form hydrophobic interactions to the side chains of these residues, or H-bond to the backbones, or weakly CH-π interactions, resulting in either favorable or undesirable protein-ligand interaction. A symbolic example is that in CYP2A6, the bulky side chain of I366 pushes the APAP methyl group toward the phenol group at the S1 binding state, resulting in significant steric repulsion; on the other hand, the same residue's backbone interacts with the ligand via a water-mediated H-bond system to help stabilize the S1r state. Third, charged or polar AA residues, usually from the highly rigid I helix, can help stabilize CYP-APAP complexes. Such examples include D295 and T303 in CYP2E1, N297 in CYP2A6, D320 in CYP1A2, and N204 and D293 in CYP2C9. APAP is more likely to form direct H-bond with these types of side chains than residues that are not polar nor charged. Finally, water molecules (especially crystallographic) play an important role in stabilizing the CYP-APAP binding complex via H-bond interactions. In CYP2E1, water-mediate H-bond patterns are found to tether APAP to the enzyme AAs,

helping to gain high affinities for the complex. Similarly, water-mediated H-bond systems are also found in other CYP-APAP complexes. In fact, water in the active site of CYP-APAP complexes may also play a catalytic role in *N*-oxidation to mediate hydrogen hopping, as reported in similar reactions[53], In addition, in the CYP2C9 SD state, a four-water-involved H-bond chain is found to fill up the space left by the absence of APAP, tethering the heme complex to the enzyme environment. Systematically analyzing these protein-ligand interactions also helps understand the different binding preferences and the product regioselectivities for the set of CYPs.

In summary, this research reveals that reliable estimation of relative free energy difference between different binding states and thorough analysis of substrate and CYP interactions are crucial to the understanding of metabolite selectivity and, in the case of APAP, drug side effect predictions. In fact, both the overlap between different sampling windows and the equilibration at each window could be well restored with fewer windows (~1/3) and shorter simulation lengths (1-2 ns). In fact, a very recent article, describing standard computing binding free energies using PMF-based approaches, reported their success with about 1 ns simulation for each window[108]. Therefore, the total number of windows and the sampling time at each window could be significantly reduced to make this approach much more computational efficient. Because a large number of sampling windows could be simulated simultaneously, this approach is highly 'parallel'. Thus, this robust free energy approach is highly suitable for high performance computing (HPC) and could play more important roles in future drug discovery and adverse effect predictions.

Conclusions

Human CYPs are responsible for most of drug metabolism. Therefore, it is critical to understand their selectivity in order to predict potential adverse drug effects. In the past, approaches based on molecular docking, MM MD, QM, and QM/MM have been reported to attempt to address the metabolism selectivity at the binding stage or the reaction stage. However, as shown in this study, neither docking nor MD simulations alone can accurately describe the equilibrium distribution of multiple drug binding

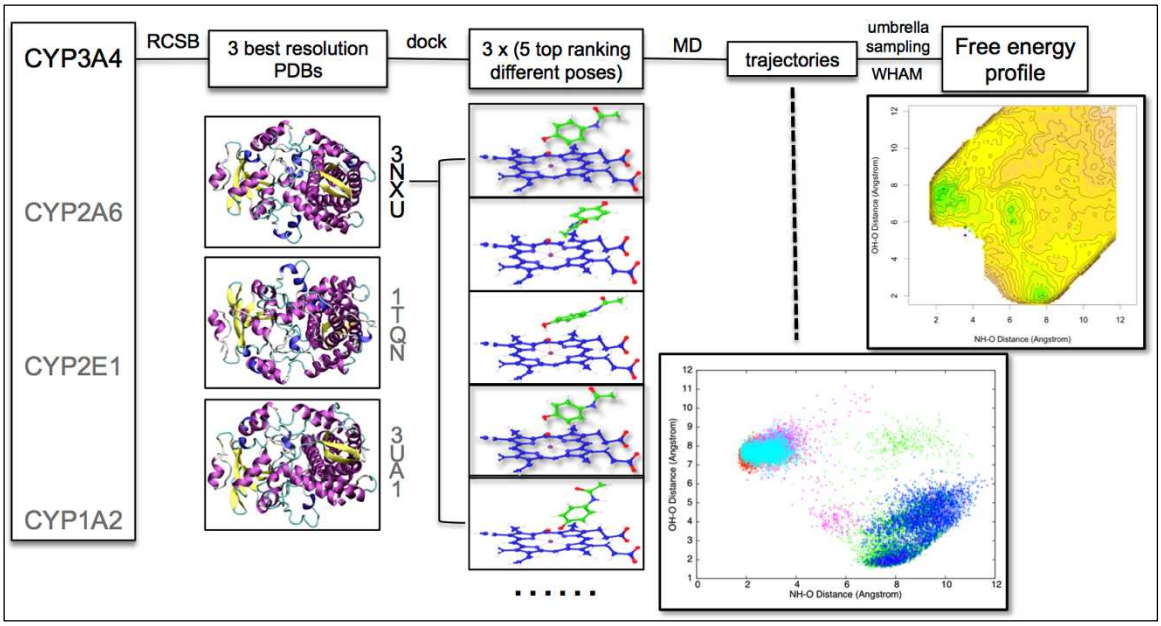
conformations in CYPs. While QM or QM/MM studies focus on the reaction step alone, they might simply ignore the energy difference prior to the chemical step or start from an inappropriate initial point. As a result, free energy simulations based on the principles of statistical mechanics and the technique of numerical simulations is necessary to fill the gap by investigating the binding landscapes. In fact, in another series of MM and QM/MM studies of the catalysis mechanism of aromatic prenyltransferase NphB conducted by the Merz lab, the authors identified that the major product is associated with a higher chemical reaction barrier than the minor product, but the experimentally measured product ratio could be reproduced by combining the energetics of both the equilibrium distribution and the chemical step[109,110]. With this in mind, we thoroughly investigated the active sites of several human CYPs and quantitatively examined the relative free energies of different binding states of APAP. At least two binding states were found for APAP in all five CYPs included in this study, with one corresponding to the NAPQI formation and another one responsible for the 3-OH-APAP formation. The relative binding free energies varied across different CYPs, from approximately 4 kcal/mol favored for the S1 state in CYP2E1 to about 2 kcal/mol more favorable of the S2 state in CYP1A2. An S3 state connecting the S1 and S2 state was identified in most CYPs except CYP2E1 (and CYP2C9 where S3 is not on the indispensible transition path between S1 and S2). A distal site was also observed in CYP1A2 though the exact function of such a remote site is not yet clear. In addition to the relative free energies between different binding states, the energy difference between the *N*-oxidation and aromatic hydroxylation chemical reaction needs to be included into consideration in order to fully understand the different APAP binding conformations that lead to metabolite selectivity as exhibited by different CYPs. Such studies require calculations at QM or QM/MM level and are currently in process in our lab. An alternative way is to estimate the energy barrier difference between two reactions based on the experimentally observed product ratio and our computed relative binding free energies, following the Curtin-Hammett principle. For CYP2A6, a product ratio of 1:3 between NAPQI and 3-OH-APAP was reported[99] and almost unperturbed by the presence or absence of other cofactors, such as cytochrome *b5*. Such a result allows us to estimate an approximately 1.7 kcal/mol free energy difference between two different chemical reactions with the *N*-

oxidation state favored. Combining this estimation with our computed relative free energy differences between different binding states, we are able to explain the different selectivities exhibited across different CYPs.

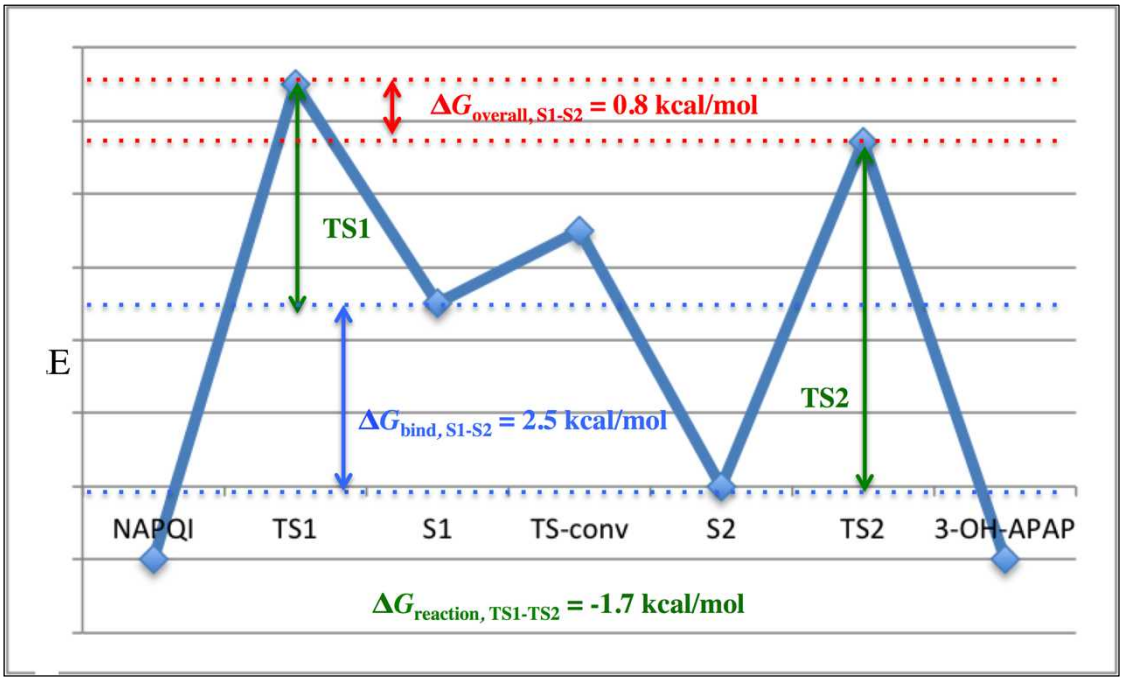
Additionally, the free energy simulations also provide us an opportunity to predict the APAP binding pose at the equilibrium states in different CYPs. Our predicted most stable APAP binding pose in CYP3A4 accurately reproduces the binding pose concluded through NMR T_1 paramagnetic relaxation experiment, including the bending between the methyl group and the rest of APAP[95]. Such an agreement further validates this free energy approach. Furthermore, our results demonstrate that the shape of the CYP binding pocket and the protein-ligand interactions are the determinants of the APAP binding preference. A remarkable example is in CYP2A6 where I366 pushes the methyl group toward the phenol group, generating strong repulsion between such two groups and making that binding state (S1) highly unstable. The binding poses identified at the basins on the free energy profile can be further used as the proper starting points for the QM or QM/MM calculations to investigate the reaction mechanism and energy barrier. Using the structures identified from our free energy approach would possibly help starting with more relevant initial structures, as a successful example following this idea has recently been reported[110,111].

Acknowledgement

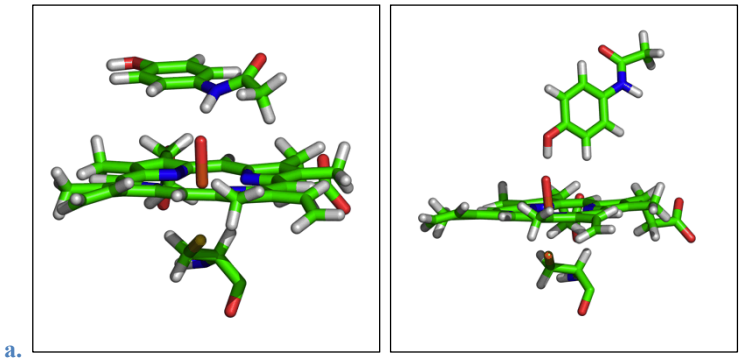
We thank Laboratory Directed Research and Development for funding, 12-SI-004. We also thank Livermore Computing for the generous computer time. This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Release number LLNL-XXXX-XXXXXX



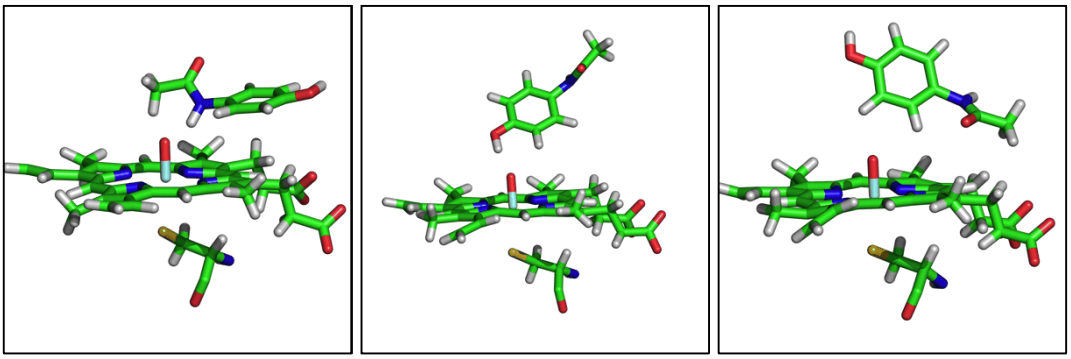
Scheme 1. Graphic illustration of the computational process of acetaminophen-CYP3A4 as an example of our research procedure: (step 1) Three high-resolution PDBs were selected for each CYP, (step 2) Five top scoring binding poses were selected for each PDB (in this example, for PDB ID 3NXU), (step 3) MD trajectories were further analyzed to extract useful information, and (step 4) free energy profile was obtained from 2-D umbrella sampling followed by WHAM analysis.



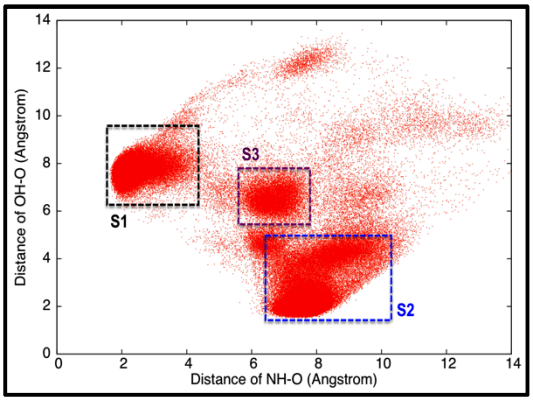
Scheme 2. Graphic illustration of the free energy chart for CYP2A6-APAP complex according to Curtin-Hammett principle.



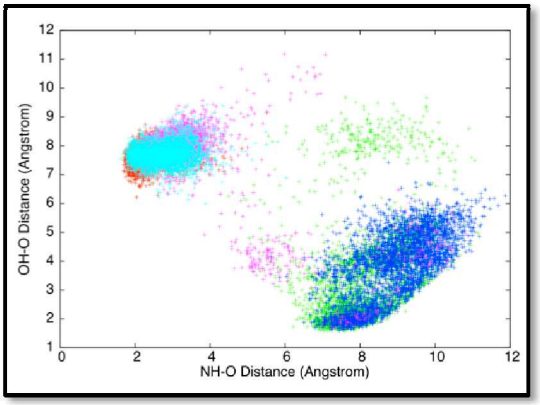
a.



b.

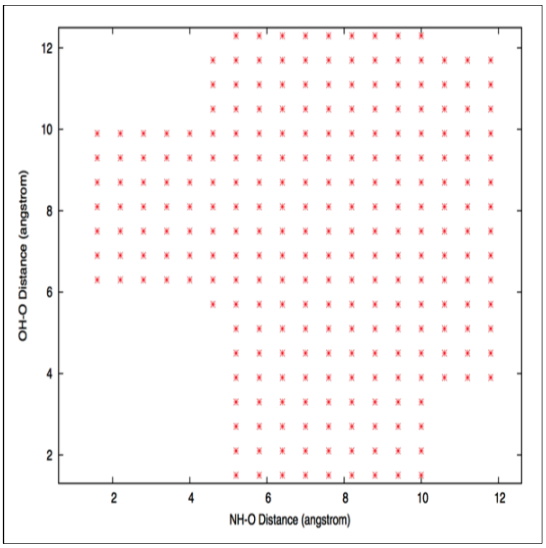


c.



d.

Figure 1. (a) CYP3A4-APAP docking poses feather close proximity of hydroxyl to heme oxygen (left) and close proximity of amide to heme oxygen (right); (b) APAP binding conformations identified from MD trajctories: close proximity between amide and heme oxygen (left), close proximity between hydroxyl and heme oxygen (middle), neither amide or hydroxyl close to heme oxygen but aromatic hydroxylation center in close proximity to heme oxygen (right). (c) distance distribution of RC1 and RC2 in all 15 MD simulations for CYP3A4, with S1, S2 and S3 state circulated in black, blue and purple boxes; and (d) distance distribution of RC1 and RC2 in five MD simulations for PDB ID 3NXU.

a. 

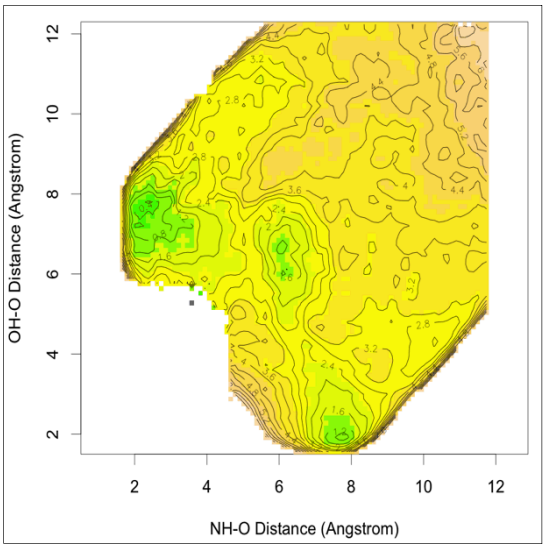
b. 

Figure 2. (a) graphic illustration of umbrella sampling windows placed along two reaction coordinates; (b) free energy profiles of APAP binding in CYP3A4. The green color represents free energy basins, yellow color means higher energy, and brown color even higher (the same below).

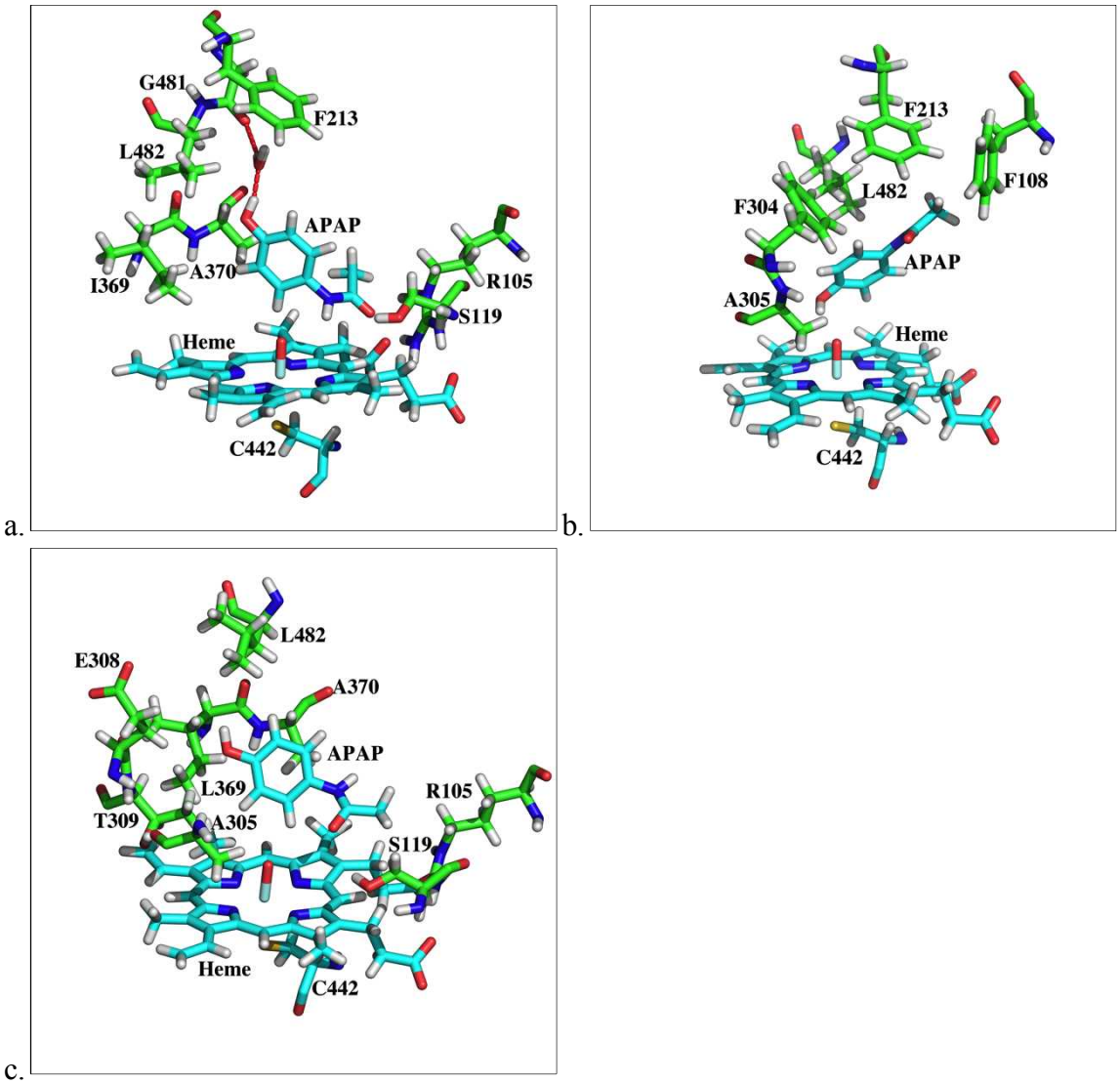


Figure 3. CYP3A4-APAP structure at the (a) S1, (b) S2 and (c) S3 state, protein side chains forming important interaction at S1 and S2 states are in the same thickness as acetaminophen and the heme center.

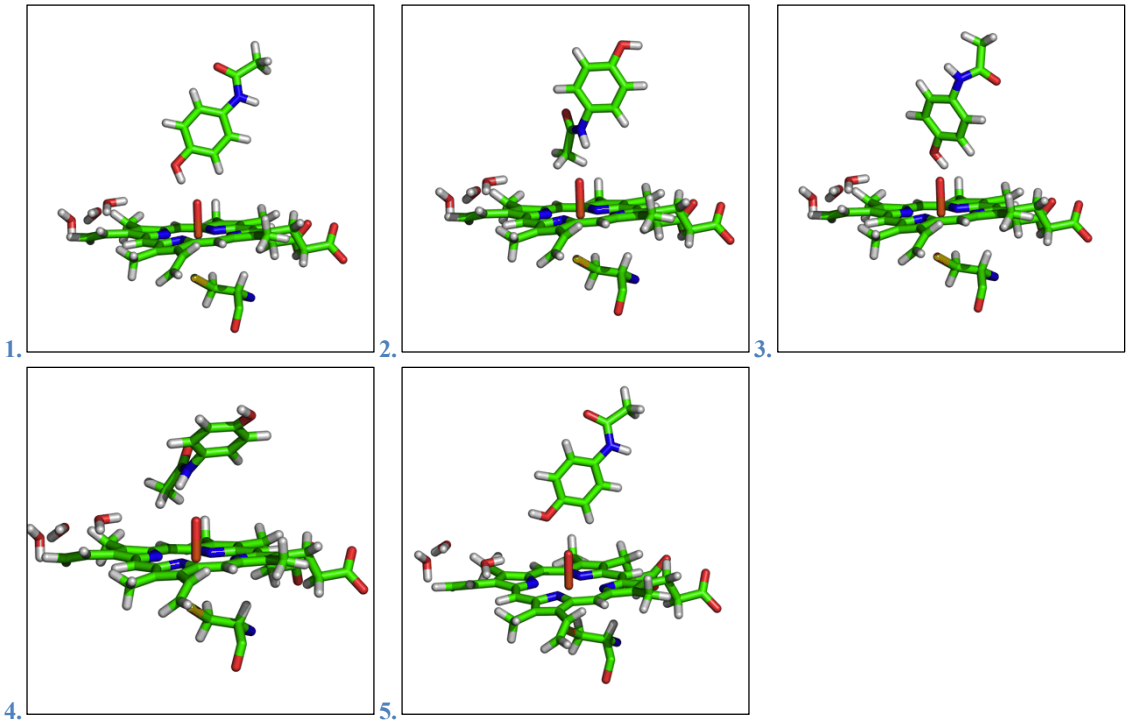


Figure 4. Top 5 (1-5) scoring binding poses from CYP2E1-APAP docking results.

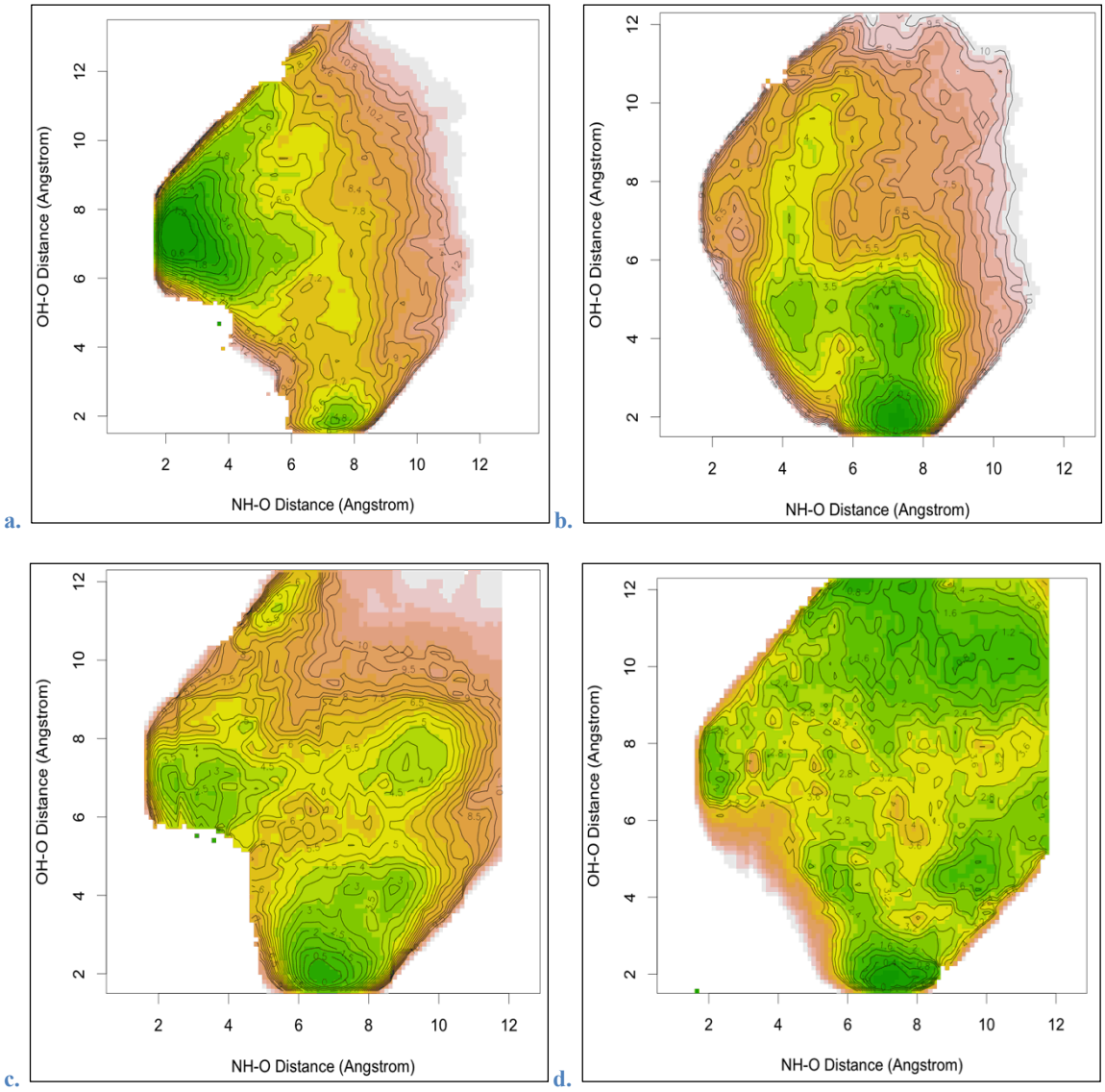
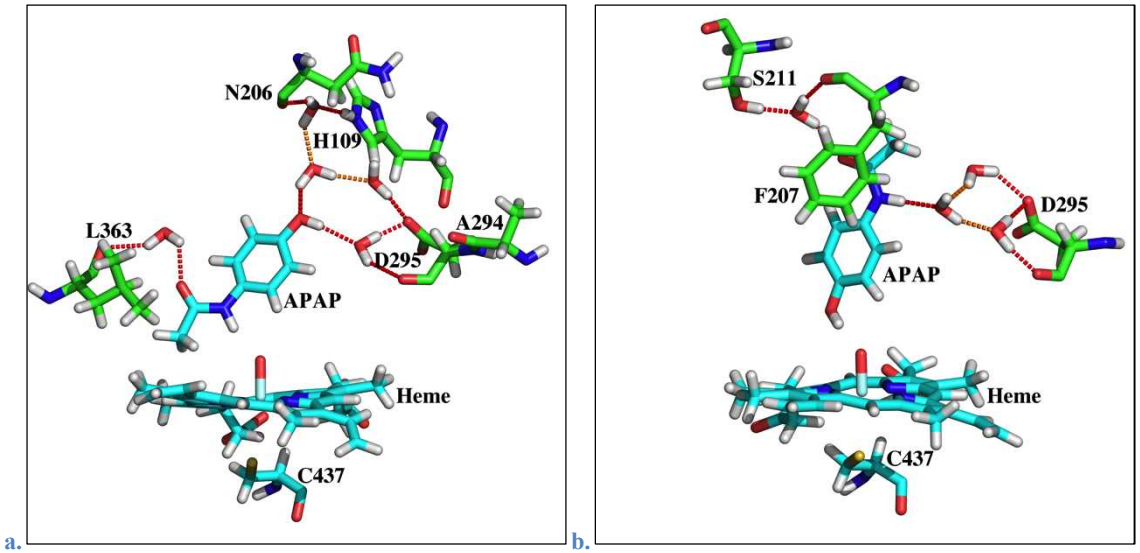


Figure 5. Free energy profiles of APAP binding in (a) CYP3A4, (b) CYP2A6, (c) CYP1A2, and (d) CYP2C9.



a. b.

Figure 6. Snapshots of the active site at the (a) S1 and (b) S2 state CYP2E1-APAP binding complex. The important water mediate hydrogen bond network are shownin dashes (same as below).

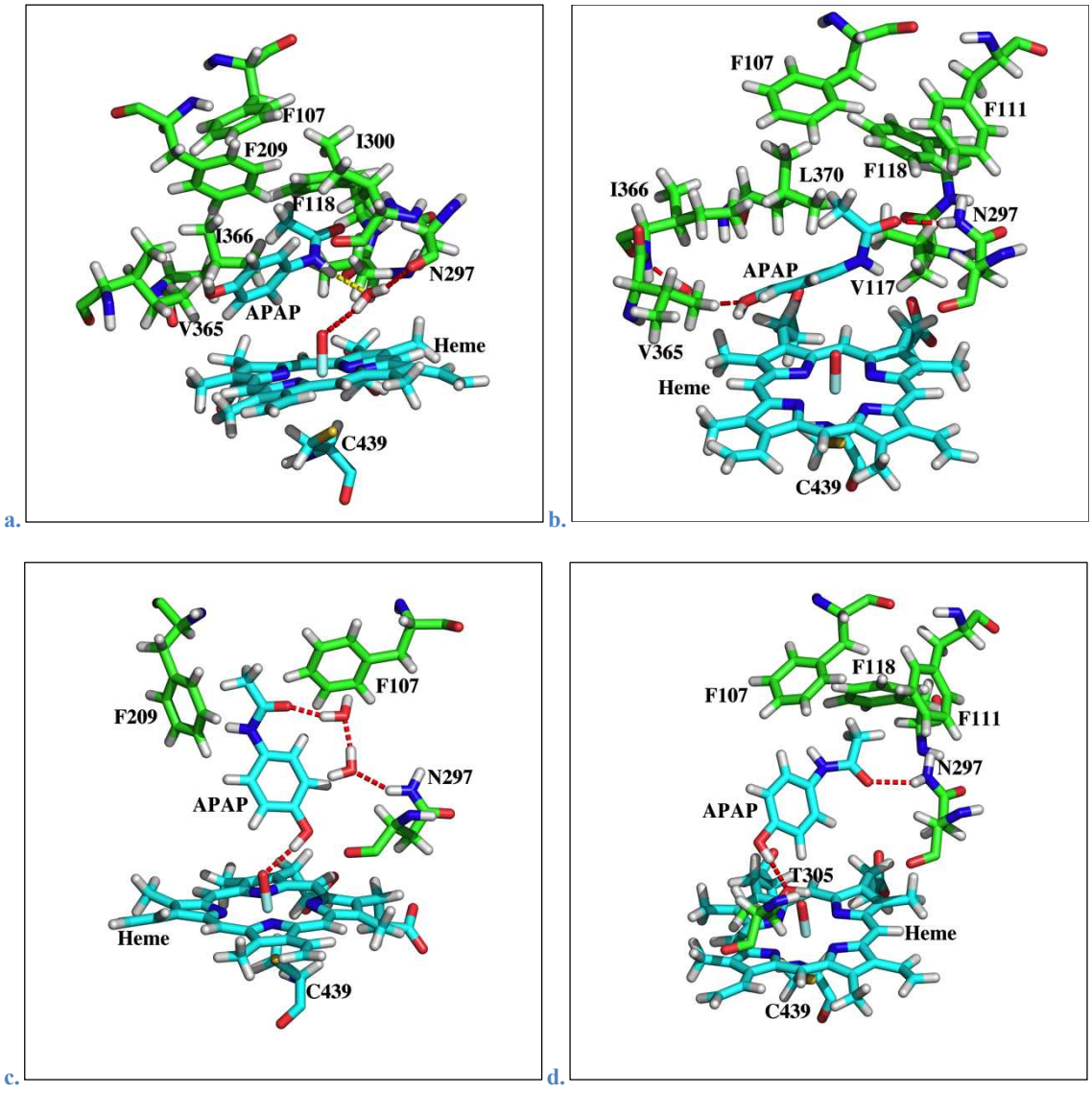


Figure 7. Structures of the active site at the (a) S1, (b) S1r, (c) S2 and (d) S3 state in the CYP2A6-APAP complex. Important hydrogen bond patterns are shown. Enzyme side chains that potentially cause steric repulsion to acetaminophen are shown in thinner sticks at S1 and S1r states, while enzyme residues possibly form CH-π interaction with acetaminophen are shown at S2 and S2 state.

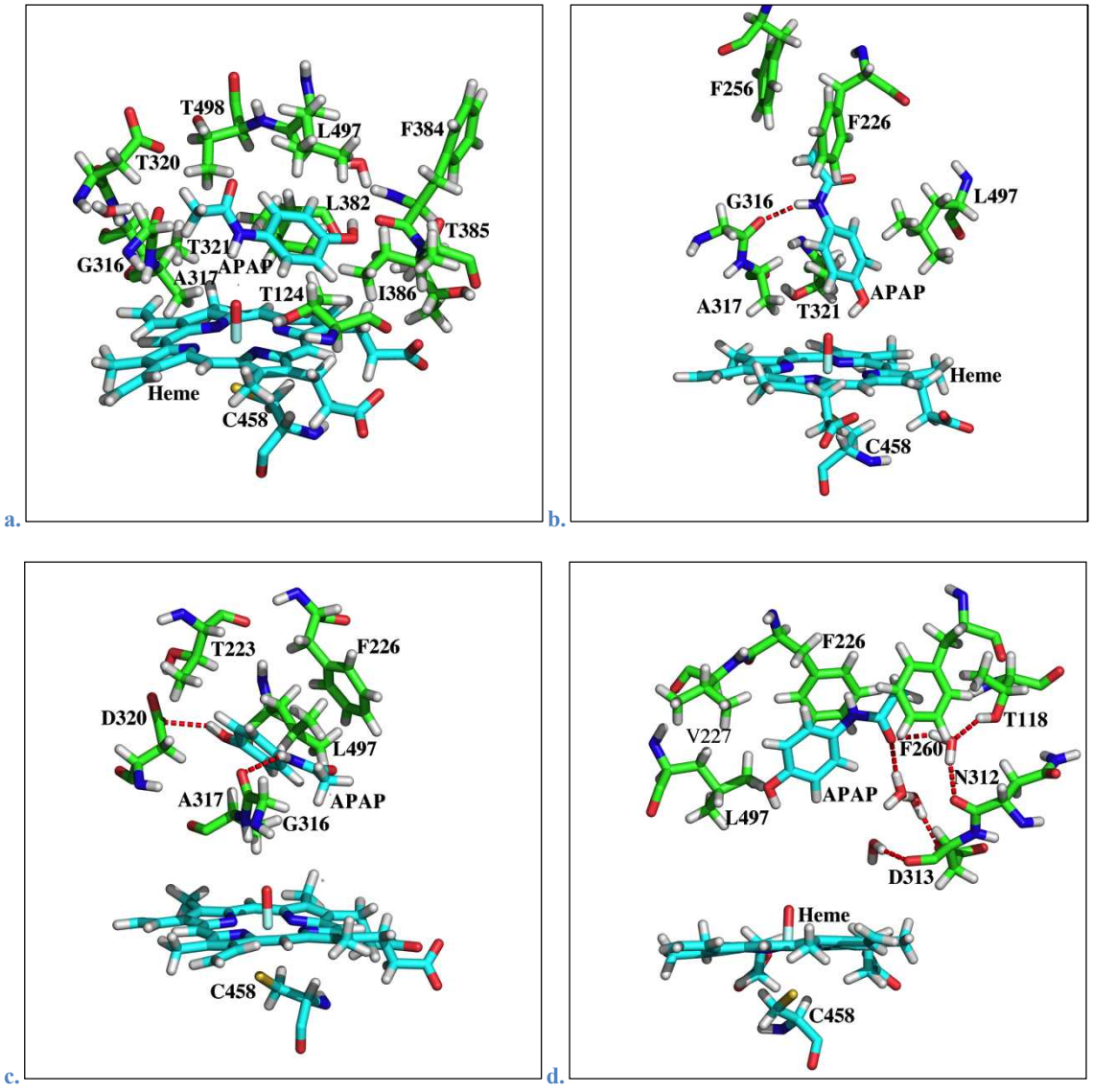


Figure 8. Active site structures of CYP1A2-acetaminophen complex at the (a) S1, (b) S2, (c) S3, and (d) SD states. Important H-bond patterns are shown.

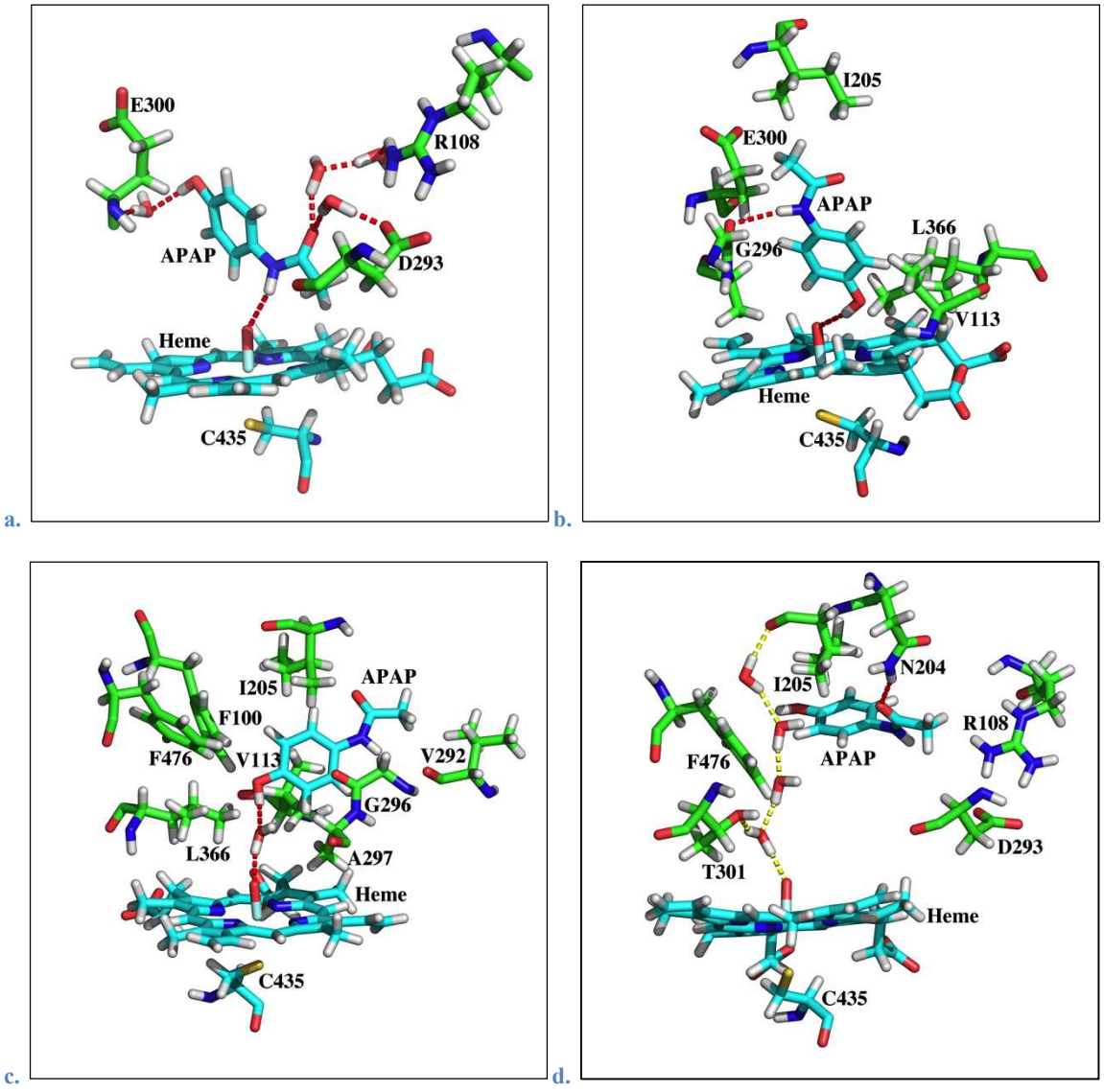


Figure 9. CYP2C9-APAP complex at the (a) S1, (b) S2, (c) s3, and (d) SD state. Important H-bond patterns are shown. In (d), the four water involved H-bond system is given in yellow dashes.

* Unless otherwise noted, the term of 'selectivity' in the article is referring to product selectivity (/metabolite selectivity/regioselectivity), meaning the preference of a CYP protein converts APAP to a specific metabolite.

(1) Danielson, P. B. *Curr Drug Metab* **2002**, *3*, 561.

(2) Guengerich, F. P. *Chem Res Toxicol* **2008**, *21*, 70.

(3) Williams, J. A.; Hyland, R.; Jones, B. C.; Smith, D. A.; Hurst, S.; Goosen, T. C.; Peterkin, V.; Koup, J. R.; Ball, S. E. *Drug Metab Dispos* **2004**, *32*, 1201.

(4) Sarich, T.; Kalhorn, T.; AlSayegh, F.; Adams, S.; Slattery, J.; Goldstein, J.; Nelson, S.; Wright, J. *Clin Pharmacol Ther* **1997**, *62*, 21.

(5) Slaughter, R. L.; Edwards, D. J. *Ann Pharmacother* **1995**, *29*, 619.

(6) Wilkinson, G. R. *New Engl J Med* **2005**, *352*, 2211.

(7) Graham, S. E.; Peterson, J. A. *Arch Biochem Biophys* **1999**, *369*, 24.

(8) Werck-Reichhart, D.; Feyereisen, R. *Genome Biology* **2000**, *1*, reviews3003.1.

(9) Omura, T.; Sato, R. *J Biol Chem* **1962**, *237*, 1375.

(10) Davydov, R.; Makris, T. M.; Kofman, V.; Werst, D. E.; Sligar, S. G.; Hoffman, B. M. *J Am Chem Soc* **2001**, *123*, 1403.

(11) Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, B. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. *Science* **2000**, *287*, 1615.

(12) Rowland, P.; Blaney, F. E.; Smyth, M. G.; Jones, J. J.; Leydon, V. R.; Oxbrow, A. K.; Lewis, C. J.; Tennant, M. G.; Modi, S.; Eggleston, D. S.; Chenery, R. J.; Bridges, A. M. *J Biol Chem* **2006**, *281*, 7614.

(13) Haines, D. C.; Tomchick, D. R.; Machius, M.; Peterson, J. A. *Biochemistry-Us* **2001**, *40*, 13456.

(14) Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. *Chem Rev* **2005**, *105*, 2253.

(15) Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. *Chem Rev* **2010**, *110*, 949.

(16) Sheng, X.; Horner, J. H.; Newcomb, M. *J Am Chem Soc* **2008**, *130*, 13310.

(17) Yang, X. A.; Huang, Y.; Crowson, M.; Li, J. R.; Maitland, M. L.; Lussier, Y. A. *J Biomed Inform* **2010**, *43*, 376.

(18) Bender, A.; Scheiber, J.; Glick, M.; Davies, J. W.; Azzaoui, K.; Hamon, J.; Urban, L.; Whitebread, S.; Jenkins, J. L. *Chemmedchem* **2007**, *2*, 861.

(19) Rydberg, P.; Gloriam, D. E.; Olsen, L. *Bioinformatics* **2010**, *26*, 2988.

(20) Rydberg, P.; Gloriam, D. E.; Zaretzki, J.; Breneman, C.; Olsen, L. *Acs Med Chem Lett* **2010**, *1*, 96.

(21) Lounkine, E.; Keiser, M. J.; Whitebread, S.; Mikhailov, D.; Hamon, J.; Jenkins, J. L.; Lavan, P.; Weber, E.; Doak, A. K.; Cote, S.; Shoichet, B. K.; Urban, L. *Nature* **2012**, *486*, 361.

(22) Matthews, E. J.; Frid, A. A. *Regul Toxicol Pharm* **2010**, *56*, 247.

(23) Frid, A. A.; Matthews, E. J. *Regul Toxicol Pharm* **2010**, *56*, 276.

(24) Hughes, J. *Pain Management: From Basics to Clinical Practice*; Elsevier Health Sciences, 2008.

(25) Koop, D. R.; Morgan, E. T.; Coon, M. J. *Fed Proc* **1983**, *42*, 910.

(26) Raucy, J. L.; Lasker, J. M.; Lieber, C. S.; Black, M. *Arch Biochem Biophys* **1989**, *271*, 270.

(27) Patten, C. J.; Thomas, P. E.; Guy, R. L.; Lee, M. J.; Gonzalez, F. J.; Guengerich, F. P.; Yang, C. S. *Chem Res Toxicol* **1993**, *6*, 511.

(28) Li, Y.; Wang, E.; Patten, C. J.; Chen, L. S.; Yang, C. S. *Drug Metab Dispos* **1994**, *22*, 566.

(29) Zhou, L. X.; Erickson, R. R.; Hardwick, J. P.; Park, S. S.; Wrighton, S. A.; Holtzman, J. L. *J Pharmacol Exp Ther* **1997**, *281*, 785.

(30) Snawder, J. E.; Roe, A. L.; Benson, R. W.; Roberts, D. W. *Biochem Bioph Res Co* **1994**, *203*, 532.

(31) Bock, K. W.; Forster, A.; Gschaidmeier, H.; Bruck, M.; Munzel, P.; Schareck, W.; Fournelgigleux, S.; Burchell, B. *Biochem Pharmacol* **1993**, *45*, 1809.

(32) Flomenbaum, N. E.; Goldfrank, L. R.; Hoffman, R. S.; Howland, M. A.; Lewin, N. A.; Nelson, L. S. *Goldfrank's toxicologic emergencies*; 8th ed.; McGraw-Hill: New York, 2006.

(33) Allameh, A.; Alikhani, N. *Toxicol in Vitro* **2002**, *16*, 637.

(34) James, O.; Lesna, M.; Roberts, S. H.; Pulman, L.; Douglas, A. P.; Smith, P. A.; Watson, A. J. *Lancet* **1975**, *2*, 579.
(35) Goldfinger, R.; Ahmed, K. S.; Pitchumoni, C. S.; Weseley, S. A. *Am J Gastroenterol* **1978**, *70*, 385.
(36) Maddrey, W. C. *J Clin Gastroenterol* **1987**, *9*, 180.
(37) Kalsi, S. S.; Wood, D. M.; Waring, W. S.; Dargan, P. I. *Open Access Emergency Medicine* **2011**, *3*, 69.
(38) Thummel, K. E.; Lee, C. A.; Kunze, K. L.; Nelson, S. D.; Slattery, J. T. *Biochem Pharmacol* **1993**, *45*, 1563.
(39) Schmidt, L. E.; Dalhoff, K. *Aliment Pharm Therap* **2003**, *18*, 979.
(40) Dong, H. J.; Haining, R. L.; Thummel, K. E.; Rettie, A. E.; Nelson, S. D. *Drug Metab Dispos* **2000**, *28*, 1397.
(41) Rendic, S. *Drug Metab Rev* **2002**, *34*, 83.
(42) Myers, T. G.; Thummel, K. E.; Kalhorn, T. F.; Nelson, S. D. *J Med Chem* **1994**, *37*, 860.
(43) Cojocaru, V.; Balali-Mood, K.; Sansom, M. S. P.; Wade, R. C. *Plos Comput Biol* **2011**, *7*.
(44) Ogliaro, F.; Cohen, S.; Filatov, M.; Harris, N.; Shaik, S. *Angew Chem Int Edit* **2000**, *39*, 3851.
(45) Ogliaro, F. O.; de Visser, S. R.; Cohen, S.; Kaneti, J.; Shaik, S. *Chembiochem* **2001**, *2*, 848.
(46) de Visser, S. P.; Shaik, S.; Sharma, P. K.; Kumar, D.; Thiel, W. *J Am Chem Soc* **2003**, *125*, 15779.
(47) Bathelt, C. M.; Mulholland, A. J.; Harvey, J. N. *J Phys Chem A* **2008**, *112*, 13149.
(48) de Visser, S. P.; Ogliaro, F.; Sharma, P. K.; Shaik, S. *J Am Chem Soc* **2002**, *124*, 11809.
(49) de Visser, S. P.; Ogliaro, F.; Sharma, P. K.; Shaik, S. *Angew Chem Int Edit* **2002**, *41*, 1947.
(50) Li, D. M.; Wang, Y.; Han, K. L. *Coordin Chem Rev* **2012**, *256*, 1137.
(51) Kumar, D.; Altun, A.; Shaik, S.; Thiel, W. *Faraday Discuss* **2011**, *148*, 373.
(52) Altarsha, M.; Benighaus, T.; Kumar, D.; Thiel, W. *J Biol Inorg Chem* **2010**, *15*, 361.
(53) Wang, Y.; Chen, H.; Makino, M.; Shiro, Y.; Nagano, S.; Asamizu, S.; Onaka, H.; Shaik, S. *J Am Chem Soc* **2009**, *131*, 6748.
(54) Altun, A.; Shaik, S.; Thiel, W. *J Am Chem Soc* **2007**, *129*, 8978.
(55) Wang, Y.; Yang, C. L.; Wang, H. M.; Han, K. L.; Shaik, S. *Chembiochem* **2007**, *8*, 277.
(56) Wang, Y.; Kumar, D.; Yang, C. L.; Han, K. L.; Shaik, S. *J Phys Chem B* **2007**, *111*, 7700.
(57) Shahrokh, K.; Orendt, A.; Yost, G. S.; Cheatham, T. E. *J Comput Chem* **2012**, *33*, 119.
(58) Olah, J.; Mulholland, A. J.; Harvey, J. N. *P Natl Acad Sci USA* **2011**, *108*, 6050.
(59) Sheng, X.; Zhang, H. M.; Im, S. C.; Horner, J. H.; Waskell, L.; Hollenberg, P. F.; Newcomb, M. *J Am Chem Soc* **2009**, *131*, 2971.
(60) Nic, M.; Jirat, J.; Kosata, B. *Crutin-Hammett principle*; IUPAC Compendium of Chemical Terminology (Online ed.), 2006.
(61) Carey, F. A.; Sundberg, R. J. *Advanced Organic Chemistry Part A Structure and Mechanisms (2nd ed.)*; Plenum Press: New York, NY., 1984.
(62) Lonsdale, R.; Olah, J.; Mulholland, A. J.; Harvey, J. N. *J Am Chem Soc* **2011**, *133*, 15464.
(63) Sansen, S.; Yano, J. K.; Reynald, R. L.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. *J Biol Chem* **2007**, *282*, 14348.
(64) Yano, J. K.; Denton, T. T.; Cerny, M. A.; Zhang, X. D.; Johnson, E. F.; Cashman, J. R. *J Med Chem* **2006**, *49*, 6987.
(65) Yano, J. K.; Hsu, M. H.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. *Nat Struct Mol Biol* **2005**, *12*, 822.
(66) Wester, M. R.; Yano, J. K.; Schoch, G. A.; Yang, C.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. *J Biol Chem* **2004**, *279*, 35630.
(67) Williams, P. A.; Cosme, J.; Ward, A.; Angova, H. C.; Vinkovic, D. M.; Jhoti, H. *Nature* **2003**, *424*, 464.
(68) Porubsky, P. R.; Meneely, K. M.; Scott, E. E. *J Biol Chem* **2008**, *283*, 33698.
(69) DeVore, N. M.; Meneely, K. M.; Bart, A. G.; Stephens, E. S.; Battaile, K. P.; Scott, E. E. *Febs J* **2012**, *279*, 1621.
(70) Sevrioukova, I. F.; Poulos, T. L. *P Natl Acad Sci USA* **2010**, *107*, 18422.

(71) Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson, E. F. *J Biol Chem* **2004**, *279*, 38091.
(72) Sevrioukova, I. F.; Poulos, T. L. *J Biol Chem* **2012**, *287*, 3510.
(73) Jacobson, M. P.; Friesner, R. A.; Xiang, Z. X.; Honig, B. *J Mol Biol* **2002**, *320*, 597.
(74) ; Schrödinger, LLC: New York, NY, 2011.
(75) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. *J Med Chem* **2004**, *47*, 1739.
(76) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. *J Med Chem* **2004**, *47*, 1750.
(77) ; Schrödinger, LLC: New York, NY, 2011.
(78) Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; V. Babin; Kollman, P. A.; University of California, San Francisco. : 2008.
(79) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.
(80) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J Comput Chem* **2004**, *25*, 1157.
(81) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J Phys Chem-Us* **1993**, *97*, 10269.
(82) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J Comput Chem* **1995**, *16*, 1357.
(83) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J Chem Phys* **1983**, *79*, 926.
(84) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, U.K, 1987.
(85) Darden, T.; York, D.; Pedersen, L. *J Chem Phys* **1993**, *98*, 10089.
(86) Uberuaga, B. P.; Anghel, M.; Voter, A. F. *J Chem Phys* **2004**, *120*, 6363.
(87) Sindhikara, D. J.; Kim, S.; Voter, A. F.; Roitberg, A. E. *Journal of Chemical Theory and Computation* **2009**, *5*, 1624.
(88) Feig, M.; Karanicolas, J.; Brooks, I. C.; MMTSB NIH Research Resource, The Scripps Research Institute: 2001.
(89) Case, D. A.; Darden, T. A.; T.E. Cheatham, I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Go□tz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A.; University of California, San Francisco: 2012.
(90) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J Comput Chem* **1992**, *13*, 1011.
(91) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J Comput Chem* **1995**, *16*, 1339.
(92) Roux, B. *Comput Phys Commun* **1995**, *91*, 275.
(93) Grossfield, A.
(94) ; R Foundation for Statistical Computing: Vienna, Austria, 2012.
(95) Cameron, M. D.; Wen, B.; Roberts, A. G.; Atkins, W. M.; Campbell, A. P.; Nelson, S. D. *Chem Res Toxicol* **2007**, *20*, 1434.
(96) Williams, P. A.; Cosme, J.; Vinkovic, D. M.; Ward, A.; Angove, H. C.; Day, P. J.; Vonrhein, C.; Tickle, I. J.; Jhoti, H. *Science* **2004**, *305*, 683.
(97) Brandl, M.; Weiss, M. S.; Jabs, A.; Suhnel, J.; Hilgenfeld, R. *J Mol Biol* **2001**, *307*, 357.
(98) Porubsky, P. R.; Battaile, K. P.; Scott, E. E. *J Biol Chem* **2010**, *285*, 22282.
(99) Chen, W. Q.; Koenigs, L. L.; Thompson, S. J.; Peter, R. M.; Pettie, A. E.; Trager, W. F.; Nelson, S. D. *Chem Res Toxicol* **1998**, *11*, 295.
(100) Carey, F. A.; Sundberg, R. J. *Advanced Organic Chemistry Part A Structure and Mechanisms*; 2nd ed.; Plenum Press: Ner York, NY, 1984.

(101) Hazai, E.; Vereczkey, L.; Monostory, K. *Biochem Bioph Res Co* **2002**, *291*, 1089.
(102) Zaher, H.; Buters, J. T. M.; Ward, J. M.; Bruno, M. K.; Lucas, A. M.; Stern, S. T.; Cohen, S. D.; Gonzalez, F. J. *Toxicol Appl Pharm* **1998**, *152*, 193.
(103) Merz, K. M. *J Am Chem Soc* **1991**, *113*, 406.
(104) Goldstein, J. A.; Demorais, S. M. F. *Pharmacogenetics* **1994**, *4*, 285.
(105) Lee, C. R.; Goldstein, J. A.; Pieper, J. A. *Pharmacogenetics* **2002**, *12*, 251.
(106) Hummel, M. A.; Locuson, C. W.; Gannett, P. M.; Rock, D. A.; Mosher, C. M.; Rettie, A. E.; Tracy, T. S. *Mol Pharmacol* **2005**, *68*, 644.
(107) Mosher, C. M.; Hummel, M. A.; Tracy, T. S.; Rettie, A. E. *Biochemistry-Us* **2008**, *47*, 11725.
(108) Gumbart, J. C.; Roux, B.; Chipot, C. *Journal of Chemical Theory and Computation* **2012**.
(109) Cui, G. L.; Li, X.; Merz, K. M. *Biochemistry-Us* **2007**, *46*, 1303.
(110) Yang, Y.; Miao, Y. P.; Wang, B.; Cui, G. L.; Merz, K. M. *Biochemistry-Us* **2012**, *51*, 2606.
(111) Yang, Y.; Wang, B.; Ucisik, M. N.; Cui, G. L.; Fierke, C. A.; Merz, K. M. *J Am Chem Soc* **2012**, *134*, 820.